The Psychosocial Consequences of Face-To-Face Versus Social Media Communication

by

Susannah Albert-Chandhok

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2021

Doctoral Committee:

Professor Ethan Kross, Chair
Associate Professor Allison Earl
Assistant Professor David Jurgens
Assistant Professor Ariana Orvell, Bryn Mawr College
Professor Oscar Ybarra

Susannah Albert-Chandhok

susac@umich.edu

ORCID iD: 0000-0002-4814-5538

Dedication

This dissertation is dedicated to my two biggest inspirations: my Mom and Dad. Thank you for teaching me how to love learning. All of who I am is because of your support.

Acknowledgements

Thank you to everyone within the University of Michigan, the Department of Psychology, and my committee for making my graduate school journey possible. In particular, I am deeply grateful for my social psychology cohort-turned-family: Clint, Imani, Martha, Nadia, and Sakura. Your support and kindness mean more to me than I can put into words.

I also have had an amazing support system in my family: Dad, Mom, Laura, Carrie, Greg, Jonathan, Mira, Sawyer, Callum, Athena, Yoshimi, and Alaska. Thank you for bringing the light into every day.

Table of Contents

List of Tables

List of Figures

Abstract

Billions of people have adopted social media as a new virtual space in which to interact with others. Yet only limited psychological research has directly compared how people communicate on social media with how they communicate in our most foundational physical space: face-to-face. In this dissertation, I propose that the *context* in which interpersonal communication takes place (i.e., face-to-face or on social media) influences how associated psychological behaviors unfold. Specifically, I find evidence that for three types of communication—self-disclosure, sharing views about controversial topics, and confrontation—the context of the interaction influences subsequent intrapersonal and interpersonal psychological outcomes. In Chapter 1, I use an experience-sampling method (N = 248) to examine how one's motivation for self-disclosure influences feelings of affective well-being and social connectedness in face-to-face versus social media contexts. In Chapter 2, I use a crowdsourcing data collection method with tweets (N = 283,587) to demonstrate that relationship closeness influences the likelihood of toxic communication during conversations about controversial topics on Twitter. In Chapter 3, I show in two studies (N = 404) that people are influenced by relationship closeness when choosing a communication context in which to engage in confrontation. In sum, by utilizing a diverse set of research methodologies, I present converging evidence in this dissertation that certain social behaviors are influenced by having taken place face-to-face or on social media.

Introduction

On December 15th, 2017, a blog post titled, "Hard Questions: Is Spending Time on Social Media Bad for Us?" went viral online. A contributor to its virality may have been its unlikely source: Facebook (Ginsberg & Burke, 2017). What precipitated Facebook—the social media company whose platform over one in seven humans use daily and whose mission is to "bring the world closer together"—to reflect on the harms of its own product?

At the end of 2017, there were many voices questioning the perceived ills of social media use. On November 9th, former Facebook President Sean Parker told *Axios* that Facebook was designed to exploit "a vulnerability in human psychology" and that using social media "changes your relationship with society, with each other" (Pandey, 2017). A day later, former Facebook Vice President Chamath Palihapitiya further emphasized this point, saying, "I think we have created tools that are ripping apart the social fabric of how society works" (Stanford University, 2017). In December, *The Atlantic* published an article titled, "Have Smartphones Destroyed a Generation?" about research findings linking an increase in social media and smartphone use in adolescent females between 2010 and 2015 with an increase in depressive symptoms and suicide-related outcomes during the same time period (Twenge, 2017; Twenge et al., 2017).

Concerns about the impact of humans using the Internet in order to fulfil social needs go back decades. One of the earliest studies on the psychology of online interactions was before the birth of social media platforms like Facebook and Twitter. In 1998, Kraut and colleagues found that increased Internet use led to decreased social involvement and increased loneliness and rates of depression (Kraut et al., 1998). The authors noted that computer-mediated communications

were limited to chat rooms where strangers superficially discussed "soap operas" and "stamp collecting," and hoped that future Internet communications might be devoted to fostering more authentic connections and social relationships.

Have social media platforms answered this call to foster deeper interpersonal connection? People use social media in staggering numbers—with over 4 billion users[1] as of 2021. And people do report using social media with the goal of communicating with others and feeling socially connected (e.g., Bazarova & Choi, 2014; Nadkarni & Hofmann, 2012; Smith, 2011; Whiting & Williams, 2013). Indeed, social connection has been pointed to as a critical building block of well-being (Baumeister & Leary, 1995; Cacioppo & Patrick, 2008), and interpersonal communication face-to-face is an effective way to feel socially connected (Cozby, 1973; Jourard, 1964; Jourard & Lasakow, 1958). But as the individuals quoted in the news articles at the beginning suggest, are the goals of using these platforms obtained? Is social media a suitable alternative to engage in effective interpersonal communication and feel socially connected?

Academic fields from psychology to information to communication to sociology have investigated this topic in order to better understand the psychology of social media use. For example, researchers have compared the effects of using different social media platforms (Bayer et al., 2016) and the effects of different types of social media usage, from more intentional and active to more passive and consuming (Verduyn et al., 2015), and the different outcomes of social media use (Hunt et al., 2018). But there is a gap in this past research that has less often been addressed – directly comparing social media as a communication context to a comparison group.

[1] https://www.statista.com/topics/1164/social-networks/

In order to address this gap, the goal of my dissertation research is to take a close look at the time people spend being social on social media and directly compare that to the time people spend being social *face-to-face*. The key research question that unites my dissertation studies is: how are social media interpersonal interactions unique as compared to interactions in face-to-face contexts?

Dissertation Overview

In the three chapters of this dissertation, I systematically examine, using diverse research methodologies, how three social behaviors—self-disclosure, discussion of controversial moralized political topics, and confrontation—play out and influence well-being and social interactions in offline and online spaces. In Chapter 1, I compare self-disclosures made face-to-face to those made on social media, taking a close look at how one's *motivations* for sharing may differentiate the outcomes of these disclosures between contexts. I also present results comparing how face-to-face versus social media self-disclosures influence affective well-being and feelings of social connection over time.

In Chapter 2, I turn from examining the individual behavior self-disclosure to the dyadic social behavior of discussing controversial moralized political topics. Specifically, I look at how the *relationship* between communication partners sharing their beliefs about divisive political topics such as abortion and gun control influences the hostility of these conversations, as they take place on Twitter. I find a robust effect that relationship closeness is *inversely* related to incivility on social media when people are discussing moralized political topics.

Finally, Chapter 3, I further examine how dyads confront one another about strongly held, but opposing, political beliefs by applying narrative analysis to examine how participants think about the decision to confront a friend or distance acquaintance either offline or online.

Chapter 1: Comparing the Psychological Consequences of Face-To-Face Versus Social Media Self-Disclosure: An Experience-Sampling Approach

Introduction

Self-disclosure, or sharing our personal thoughts and feelings with other people, is a fundamental component of human sociality (Cozby, 1973; Jourard, 1964; Jourard & Lasakow, 1958). These acts of revealing self-related information to others are the building blocks of forming new relationships with other people, as well as maintaining current relationships (Canevello & Crocker, 2010; Collins & Miller, 1994). In many face-to-face contexts, self-disclosure has been shown to increase positive affect (Diener & Seligman, 2002; Hawkley & Cacioppo, 2010; Mehl et al., 2010; Watson et al., 1992) and foster feelings of social connection (Altman & Taylor, 1973; Canevello & Crocker, 2010; Collins & Miller, 1994; Cozby, 1973; Laurenceau et al., 1998).

In general, we are motivated to share our emotions and personal experiences with others to achieve many, and often overlapping, goals, from social (e.g., building intimate connections, receiving validation and support) to cognitive (e.g., creating meaning from events) (Butzel & Ryan, 1997; Rimé, 2009; Rimé et al., 1991) to emotional (e.g., interpersonal emotion regulation) (Zaki & Williams, 2013). Neurological evidence suggests that we are motivated to self-disclose because it feels good; research shows that when people self-disclose, there is increased activation in areas of the brain associated with reward, like the ventral striatum (Tamir & Mitchell, 2012). Given how central self-disclosure is to healthy psychological functioning, how many motivations and goals it can serve, and how rewarding it is, it's unsurprising that self-disclosure is frequent—

we spend 30-40% of our time during face-to-face conversations sharing our inner thoughts and feelings with others (Dunbar et al., 1997).

For thousands of years, humans have continually developed new technologies to amplify our capacity to self-disclose to others – from the invention of writing to the printing press to the telephone. But since the early 2000s, we've seen a new technology erupt in popularity, one that gives us a remarkable ability to easily and efficiently broadcast what we think and feel to a vast audience: social media.

Over half the world's population, or 3.96 billion people, uses some form of social media (Clement, 2020) and, to use Facebook as an example, there are 3.3 million new posts from users every minute (Allen, 2017). In fact, two of the most popular social media platforms with millions of users—Facebook and Twitter—explicitly encourage people to self-disclose immediately upon logging into these sites. At the top of their respective homepages, Facebook asks, "What's on your mind?" and Twitter inquires, "What's happening?" with a text box below for people to write their responses (Figure 1). And indeed, social media users frequently respond to these prompts. Self-disclosure is as fundamental a part of online social behavior as it is offline—as much as 80% of direct communications on social media involve sharing one's own thoughts and feelings (Carr et al., 2012; Moreno et al., 2011; Naaman et al., 2010).

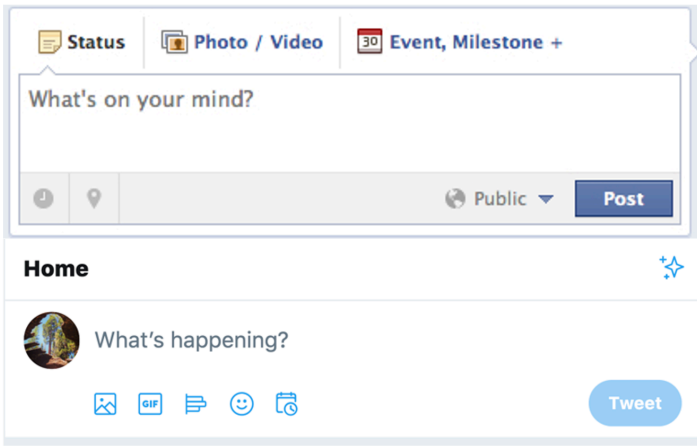

Figure 1. Screenshots of the home page when a user logs onto Facebook (top) and Twitter (bottom), showing text boxes that prompt self-disclosure.

A recent but growing body of work has begun to examine self-disclosure in the context of social media interactions. However, this work has produced mixed findings as to the costs and benefits of social media self-disclosure on well-being and social connectedness. On the one hand, some research has found that sharing status updates on Facebook is associated with increased feelings of connectedness to other social media users (Leimeister et al., 2011; N. Park et al., 2011) and less loneliness generally (Deters & Mehl, 2013; Lee et al., 2013). On the other hand, some research indicates that posting status updates is associated with decreases in self-reported mental health over time (Shakya & Christakis, 2017). Experimental work has found no significant effect in either direction of social media self-disclosure on affective well-being over time (Verduyn et al., 2015).

Despite some ambiguity about the emotional *consequences* of self-disclosure online, the literature does show that people are indeed *motivated* to self-disclose and share personal experiences on social media to achieve a range of different social (e.g., relationship development, validation), cognitive (e.g., creating a personal identity, learning from others), and emotional (e.g., interpersonal emotion regulation) goals (Bazarova et al., 2015; Bazarova &

Choi, 2014; M. Choi & Toma, 2014; Y. H. Choi & Bazarova, 2015; Nadkarni & Hofmann, 2012; Smith, 2011; Whiting & Williams, 2013), similar to offline self-disclosure.

With prior research, we can begin to indirectly compare self-disclosure in the offline and online worlds. But despite robust literatures emerging individually in each domain, no work that I am aware of has examined how these contexts *compare* in terms of their effects on affective well-being and social connection. I take as my starting point that in order to understand the similarities and differences between offline and online self-disclosure, we need to examine the motivations people have for self-disclosure and how people may differentially endorse these motivations when reflecting on self-disclosure offline versus online. Specifically, in this paper, I examine how four different motivations for self-disclosure interact with the context of their disclosures (i.e., face-to-face versus on social media) to influence how the self-discloser feels, both subjectively and in terms of how connected they feel with other people.

Self-Disclosure: A fundamental drive that satisfies multiple goals

Humans have a fundamental, evolutionarily-based drive to create and maintain social bonds (Baumeister & Leary, 1995; Dunbar, 1998). Indeed, building meaningful social ties with other people is associated with improved psychical and psychological health (Baumeister & Leary, 1995; Cacioppo & Patrick, 2008; Hawkley & Cacioppo, 2010). Having meaningful social interactions increases our feelings of positive affect in the short and long-term (Ryan & Deci, 2001; Watson et al., 1992). In general, happier people have many satisfying social relationships and more substantive social interactions than less happy people (Diener & Seligman, 2002; Mehl et al., 2010).

But what makes a social interaction "meaningful" or "substantive"? In the offline world, such social interchanges are characterized by higher levels of reciprocal self-disclosures (Cozby,

1973; Jourard & Lasakow, 1958). Self-disclosure increases feelings of intimacy between conversational partners (Reis & Shaver, 1988), and reciprocal self-disclosures during face-to-face social interactions engender feelings of liking and in turn help people develop meaningful social connections with one another (Collins & Miller, 1994).

While the distal causes for self-disclosure may have deep evolutionary roots, there are also many more proximate motivations people have for self-disclosing face-to-face, and in all likelihood on social media as well. However, it is difficult in a single study or even a set of studies to test out all possible motivations people may have for self-disclosure. Thus, in this paper, I will focus on examining four specific possible motivations people may have for self-disclosure offline and online: (1) to feel good, (2) to connect with others, (3) to manage impressions, and (4) without any deliberate motivation or desired goal. This last motivation, to share without a thought of an intended end-state, is the more impulsive counterpart to the first three, more deliberate, motivations. While not exhaustive, these four motivations are ones that have been explored in past research and that should be theoretically relevant for comparing the psychological consequences of self-disclosure in offline versus online contexts.

*Motivation #1: Self-disclosing **to feel good**.* First, people may be hedonically motivated to share because they predict self-disclosure will feel good, i.e., increase subjective positive affect over time. In offline contexts, research has generally shown this prediction to be accurate. Self-disclosures made face-to-face feel good because they do indeed increase positive affect and decrease negative affect over time (Kahn et al., 2001; Vittengl & Holt, 2000). Furthermore, self-disclosure activates brain regions associated with reward (Tamir & Mitchell, 2012). Interestingly, this elevation of positive affect is not contingent upon having an audience—people feel good even when privately reflecting on self-related thoughts and feelings—however,

disclosing to an audience more strongly activates feelings of reward (Tamir & Mitchell, 2012; see Figure 2). This suggests that responsiveness from another person during a face-to-face self-disclosure can amplify positive affect, but it's not essential.

Face-to-face self-disclosure does not always beget positive affect. Individuals with lower levels of self-esteem may not benefit from disclosure for fear of negative appraisal and rejection (Cameron et al., 2009; Wood & Forest, 2016). Relatedly, co-rumination, or excessively discussing personal thoughts and feelings within a dyad, has been shown to lead to negative outcomes such as increased emotional distress and anxiety (Calmes & Roberts, 2008; Rose, 2002). However, in spite of these defined boundary conditions, research has consistently shown the overall impact of face-to-face self-disclosure to be positive.

Conversely, findings less consistently show that social media self-disclosures increase positive affect over time. It is the case that people report being motivated to post on social media because they *believe* doing so will have hedonic value (Park et al., 2009; van der Heijden, 2004; Whiting & Williams, 2013). And there is evidence that receiving social support in the form of Facebook "Likes" on status updates does feel good and activate neural reward circuitry in the moment (Sherman et al., 2016).

However, other work suggests that self-disclosure on social media does not significantly increase subjective well-being over time (Shakya & Christakis, 2017; Verduyn et al., 2015). For example, over the course of a series of studies, Kross and colleagues found that active social media use (not exclusively self-disclosure, but also engagement on social media in the form of "Likes" and comments) does not significantly predict changes in affective well-being over time (Kross et al., 2013, 2018; Verduyn et al., 2015). In one longitudinal study, other researchers found that posting more status updates (again, which may not exclusively be acts of self-

disclosure, but does include it) was associated with a decrease in mental health over time (Shakya & Christakis, 2017). Finally, in one cross-sectional study, the authors found that self-disclosure on social media has no *direct* effect on well-being; instead, the effect is fully mediated by perceptions of receiving high social support to those disclosures (Lee et al., 2013).

Based on this prior research, I predict that while rates of endorsing self-disclosure as a means to feel good is similar offline and online, sharing to feel good offline leads to increases in positive affect over time, while sharing with the same motivation on social media has non-significant effects. I base this prediction on prior literature, which consistently shows self-disclosures made face-to-face increase positive affect over time (when not looking at specific populations like individuals with low self-esteem or specific processes like co-rumination). However, the same relationship has not been consistently found in research on self-disclosure on social media. I would predict that the desire to share in order to feel good alone is not sufficient, but that the desire to receive social support is essential to self-disclosures on social media feeling good. While sharing face-to-face may have intrinsic intrapersonal value, I think sharing on social media does not have this same intrinsic value, and only comes from interpersonal connection. This is based on literature suggesting that increases in positive affect after sharing online may come fully from perceptions of having a supportive audience and feelings of social connection (Lee et al., 2013). I will explore this motivation of sharing to make social connections in the next section.

Motivation #2: Self-disclosure as a means ***to socially connect.*** Self-disclosure, in offline and online contexts, has been shown in multiple studies to be an effective strategy to increase feelings of social connection with others. In offline contexts, many decades of research has shown that self-disclosure fosters the formation and maintenance of social connections via

increasing intimacy, vulnerability, and authenticity between conversational partners (Altman & Taylor, 1973; Bruk et al., 2018; Canevello & Crocker, 2010; Jourard & Lasakow, 1958; Laurenceau et al., 1998; Reis & Shaver, 1988; Sprecher et al., 2013).

Many people endorse a need to belong and to build social connections as their motivations for using social media (Nadkarni & Hofmann, 2012; Smith, 2011; Whiting & Williams, 2013). The mission of the social media site Facebook is "to build community and bring the world closer together," so it follows that this motivation might be a central reason for many social media users to use the site (Facebook, 2019).

Research suggests this motivation is being satisfied. On social media, self-disclosure has been linked to outcomes of feeling more socially connected and less lonely over time (Burke & Kraut, 2014; Deters & Mehl, 2013; Grieve et al., 2013; Utz, 2015). Sharing one's thoughts and feelings on social media has also been linked to related increases in feelings of interpersonal trust and social capital (Ellison et al., 2007; Valenzuela et al., 2009). Additionally, users are more likely to self-disclose on social media if they perceive their networks to be more responsive (Walsh et al., 2020). One article proposes that self-disclosure online enhances relationship quality above and beyond face-to-face self-disclosure because there are fewer inhibitions when disclosing online (Valkenburg & Peter, 2009). Lastly, a recent framework proposed by Luo and Hancock about how self-disclosure on social media may lead to changes in well-being largely depends on interpersonal mechanisms, such as perceived connectedness, social support, and social capitalization (Luo & Hancock, 2020)

Based on this prior research, I predict that rates of endorsing self-disclosure as a means to socially connect are similar offline and online, and also that sharing to connect leads to increases in positive affect and social connectedness over time, both offline and online. I would expect

stronger main effects of sharing to connect in the face-to-face condition. However, I would also expect to see moderation by perceived audience responsiveness in the social media condition, such that if audiences are perceived to be more responsive, the link between the motivation to share to connect and affective well-being and social connectedness outcomes is stronger.

Motivation #3: Self-disclosure as a means ***to manage impressions.*** The act of self-disclosure gives us the chance to deliberately attempt to influence how others form impressions about us (Goffman, 1959). Research on strategic self-disclosure in-person often considers the intrapersonal consequences of self-enhancement, or deliberate attempts to present ourselves in a more positive light, (e.g., Brown, 1986; Paulhus, 1998). Strategic self-enhancement during self-disclosure does feel good to the sharer; however, the effects on the listener may be the opposite (Paulhus, 1998; Schlenker & Leary, 1982; Taylor & Armor, 1996).

In multiple articles on computer-mediated self-disclosure, Walther argues that the desire to selectively self-present ourselves to manage impressions motivates online self-disclosure even more so than face-to-face, which he terms the hyperpersonal model of computer-mediated communication (Walther, 2007; Walther et al., 2015). Research on dating profiles shows that users take advantage of certain online affordances (such as more time to construct and edit a disclosure) to make more positive self-presentations (Toma et al., 2008). And social media users in general report being motivated to use social media in order to have more control over self-presentation (Bazarova & Choi, 2014; Nadkarni & Hofmann, 2012).

Does this motivation to self-disclose in order to manage impressions lead to the discloser feeling better and/or more socially connected over time on social media? In one study, researchers found that using social media and viewing one's own Facebook profile can provide a source of self-affirmation and improve wellbeing (although one's Facebook profile may not be

exclusively comprised of self-disclosures) (Toma & Hancock, 2013). However, in focus groups with Facebook users, researchers found that the desire to post during a positive life event (like while on a vacation or when hanging out with friends) in order to self-enhance on social media can be exhausting and reduce enjoyment of the event itself (Fox & Moreland, 2015). Overall, it's not clear based on experimental evidence if this motivation to manage impressions is more strongly endorsed for online versus offline self-disclosures, or if it leads to differential outcomes.

I predict that rates of endorsing self-disclosure as a means to manage impressions are significantly higher on social media than face-to-face, based on Walther's hyperpersonal model (Walther, 2007). However, I expect that sharing to manage impressions face-to-face leads to stronger increases in affective well-being and social connection than on social media, given the more consistent literature that sharing to self-enhance feels good face-to-face (e.g., Paulhus, 1998), along with more anecdotal evidence that doing the same on social media may have negative consequences (e.g., Fox & Moreland, 2015).

Motivation #4: Self-disclosure ***without thought.*** Finally, self-disclosure might occur without a deliberate motivation or without thought at all. Indeed, as informed by dual-process theories, researchers have found that many psychological processes can be broadly broken down into two paths: one where processes are relatively fast, less thoughtful, and "hot" or one where processes are relatively slower, more thoughtful, and "cool" (Chaiken & Trope, 1999). This motivation I am proposing—self-disclosure without thought—is a "hot" process whereas the prior three motivations I described—self-disclosure to feel good, to connect, and to manage impressions—are "cool" and more deliberate.

Although this idea of "hot" self-disclosures has been less examined in face-to-face contexts as a motivation for self-disclosure, it has been examined as a motivation for sharing on

social media. Online, people have restricted awareness of themselves and feedback from others (Joinson, 2001; Kiesler et al., 1984; Wang et al., 2011). This lack of feedback, which exists in face-to-face social interactions, can break down barriers to self-expression and fuel impulsive posting without any deliberate motivation, despite people reporting being concerned about privacy and oversharing (Acquisti & Gross, 2006). Such oversharing without deliberate motivation has been linked to negative outcomes such as regret and embarrassment (Agger, 2015; Wang et al., 2011) as well as problematic and risky social media use (e.g., sharing too much personal information, swearing, etc.).

Based on this prior work, I predict that rates of endorsing self-disclosure without thought are significantly higher on social media than face-to-face, but that endorsing this motivation leads to decreases in positive affect and social connectedness over time online. Offline, I believe a more exploratory perspective is warranted in regard to how this motivation for sharing impacts well-being.

In summary, these four motivations for self-disclosure—to connect, to feel good, to manage impressions, and with no thought—are not exhaustive, but are fundamental and have been previously examined in studies in offline and online contexts in multiple studies. Thus, the goal of the current investigation is to compare face-to-face and social media self-disclosures directly, through the lens of what motivates these self-disclosures and how these motivations influence how sharing impacts well-being and feelings of social connection over time.

To investigate this topic, I employed an experience-sampling method approach, which allows the reliable evaluation of the nuances of people's moment-to-moment experiences in daily life (Bolger et al., 2003; Csikszentmihalyi & Larson, 2014). With this method, we collected in-

vivo data about people's self-disclosures face-to-face and on social media, then used multilevel modeling to uncover how self-expression motivation influences well-being over time.

Method

Overview. This study consisted of three phrases. In Phase 1, participants completed a battery of baseline questionnaires in the lab and reviewed procedures for the experience-sampling method portion of the study (i.e., Phase 2) with a trained experimenter. During Phase 2, participants were text-messaged five times per day for a period of 14-21 days. From September 2017 - April 2018, we collected data over 14 days, and from September 2018 - December 2018, we collected data over 21 days. The decision to increase the study period after April 2018 was made in order to collect more individual occurrences of social media posting for each participant. Membership in 14- or 21-day sample did not moderate results. Each text message contained a link to a Qualtrics online survey, which asked participants to answer questions about their affective well-being, feelings of social connectedness, and experiences of self-disclosure. In Phase 3, participants returned to the lab where they were debriefed.

Participants. Two hundred and forty-eight University of Michigan undergraduates (M_{age} = 18.71, SD_{age} =1.22; 154 females; 55% White, 23% Asian, 8% African American, and 15% multiracial or other) were recruited to participate in this study in exchange for course credit (up to two hours of credit for their participation). To qualify for the study, participants had to be over 18 years of age, own a smartphone, and regularly use Facebook and Twitter. The University of Michigan Institutional Review Board approved this study. Informed written consent was obtained from all participants prior to participation. Our target sample size was informed by running as many participants as possible given our university subject pool's available resources for an academic year, beginning in the Fall 2017 semester and concluding at the end of the Fall

2018 semester (excluding summer semester). Participants who completed at least 90% of the texts were also entered into a lottery to receive a $50 Amazon gift card.

Phase 1. ***Baseline Measures.*** Participants completed a set of four questionnaires: the Extraversion subscale of the Big Five Inventory (α = 0.88, *M* = 3.35, *SD* = 0.79; John, Naumann, & Soto, 2008), the Rosenberg Self-Esteem Scale (α = 0.90, *M* = 3.78, *SD* = 0.72; Rosenberg, 1965), the Need to Belong Scale (α = 0.83, *M* = 3.46, *SD* = 0.66; Leary, Kelly, Cottrell, & Schreindorfer, 2013), and the Trait Self-Control Scale (α = 0.84, *M* = 3.17, *SD* = 0.63; Tangney, Baumeister, & Boone, 2004). None of these baselines measures moderated results. Participants also answered a series of standard demographic questions about their gender, age, and ethnicity.

Phase 2. ***Experience-Sampling Phase.***

Overview. Using experience-sampling methods (Bolger et al., 2003), participants were text-messaged five times per day between 10 a.m. and 10 p.m. for between 14 to 21 consecutive days. Text messages were sent at random times within five 168-minute intervals each day. Each text message contained a link to a brief online survey hosted by Qualtrics. The survey either asked participants about their most recent *face-to-face* self-disclosures (i.e., sharing what's on your mind face-to-face with another person) or their most recent *social media* self-disclosures (i.e., sharing what's on your mind on Facebook and/or with a post). Participants were randomly assigned to receive either the face-to-face or social media survey during the first half of the Phase 2 period, and then received the other survey during the second half (there were no significant order effects). Each survey was identical except where indicated below. In total, we collected 13,060 data points from 219 participants.

*Attrition and compliance***.** Participants responded to an average of 68.10% of text messages. Following prior work (Kross et al., 2013; Moberly & Watkins, 2008; Verduyn et al.,

2015), we excluded participants who responded to fewer than 33% of the surveys on a priori grounds. This exclusion criteria applied to 29 participants, leaving us with 219 participants for the subsequent analyses.

Items. Each survey asked participants six questions (Table 1). First, participants were asked, in a fixed order, "How happy are you feeling right now?" (*not at all* [0] to *very much* [10]; $M = 6.01$, $SD = 2.22$) and "How socially connected are you feeling right now?" (*not at all* [0] to *very much* [10]; $M = 5.29$, $SD = 2.53$).

Next, participants were asked if they had self-disclosed since the last text message that they received from us, either in a face-to-face context ($n_{face\text{-}to\text{-}face} = 2{,}948$) or in social media context, specifically on Facebook and/or Twitter ($n_{socialmedia} = 594$). This discrepancy between number of self-disclosures, with participants reporting over four times as many face-to-face self-disclosures (2,948) as compared to social media self-disclosures (594), is aligned with previous research that shows direct face-to-face communication is significantly more frequent than direct social media communication (Verduyn et al., 2015; see Figure 2). Participants who had not self-disclosed since the last text message (n = 9,482) were subsequently asked to complete filler questions about their most recent meal.

Participants who reported that they had self-disclosed since the last survey were subsequently asked to describe this self-disclosure based on two features: first, participants were asked, "How positive was the thing you shared face-to-face (posted on social media)?" (*not at all* [0] to *very much* [10]; $M = 6.25$, $SD = 2.48$) and second, they were asked "How responsive and thoughtful were people when you shared face-to-face (posted on social media)?" (*not at all* [0] to *very much* [10]; $M = 7.08$, $SD = 2.08$). We varied the wording of the face-to-face versus social

media questions to account for differences in how self-disclosure is conceptualized in these different communication contexts (i.e., face-to-face *share* versus social media *post*).

Finally, participants were asked to rate the extent to which four possible motivations drove their most recent self-disclosure. These four motivations were not exclusive—in other words, participants could endorse multiple motivations. The four motivations were (1) social connection (i.e., "I wanted to connect with another person") [*"social connection"]*, (2) pleasantness (i.e., "I thought it would feel good") [*"feel good"]*, (3) impression management (i.e., "I wanted another person to think highly of me") [*"to make a good impression"]*, and (4) impulsiveness (i.e., "I don't know, I didn't think twice about it" [*"with no thought"])*. For each of these four motivations, participants were asked to indicate, "Rate how much you agree with each reason" (*strongly disagree* [0] to *strongly agree* [10] for each motivation).

Variable	Question
Affective well-being	How happy are you feeling right now?
Social connection	How socially connected are you feeling right now?
Number of self-disclosures	*(Face-to-face)* Since the last time we texted, how many times did you share what's on your mind, in a face-to-face conversation? *(Social media)* Since the last time we texted, how many times did you share what's on your mind, with a Facebook or Twitter post?
Perceived self-disclosure positivity	*(Face-to-face)* How positive was the thing you shared? *(Social media)* How positive was your post?
Perceived self-disclosure audience responsiveness	*(Face-to-face)* How responsive and thoughtful were people when you shared? *(Social media)* How responsive and thoughtful were people to your post?

Self-disclosure motivation	*(Face-to-face)* Why did you share what's on your mind? Rate how much you agree with each reason. *(Social media)* Why did you make this post? Rate how much you agree with each reason. Reasons: 1. *[Social connection]* I wanted to connect with another person. 2. *[Feel good]* I thought it would feel good. 3. *[To make a good impression]* I wanted to make another person think highly of me. 4. *[With no thought]* I don't know, I didn't think twice about it.

Table 1. Text-message survey questions from the experience-sampling phase. Participants rated affective well-being, social connection, disclosure positivity, and audience responsiveness from *0 = Not at all* to *10 = Very much*; participants rated self-disclosure motivation from *0 = Strongly disagree* to *10 = Strongly agree*.

Phase 3. After the experience-sampling phase of the study was complete, participants returned to the lab where they were debriefed by a trained experimenter and compensated for their participation.

Results

Part I: The four motivations modeled as separate predictors

Analysis Overview. Following prior work with similar experimental designs (Kross et al., 2013, 2018; Verduyn et al., 2015), I examined the relationship between the four self-disclosure motivations, self-disclosure context (i.e., face-to-face or on social media), affective well-being, and social connection using multilevel modeling to account for the nested data structure (i.e., assessment time points nested within participants). In Part I of the results, I examine the four motivations as separate predictors; in Part II of the results, I examine if dimensionality reduction is appropriate to collapse the number of predictors.

More specifically, in this section, I examined if T_2 affect ("How happy are you feeling right now?") or T_2 feelings of social connection ("How socially connected are you feeling right now?") were predicted by $T_{1\text{-}2}$ motivation for self-disclosure, in either a face-to-face or social media context, controlling for T_1 affect or feelings of social connection at level-1 of the model. Note that although this analysis assesses motivation for a self-disclosure at T_2, this research question refers participants to reflect on their motivation *between* T_1 and T_2 (hence the notation $T_{1\text{-}2}$). Therefore, this analysis allowed us to examine whether self-disclosure motivation during the time period separating T_1 and T_2 predicted a *change* in affect or feelings of social connection over time. Following prior work (Koval et al., 2012), between-day lags were excluded from the lagged analyses (i.e., participants first ratings in the morning were not predicted by their last ratings on the previous day).

All level-1 covariates were group-mean centered (or centered within clusters), following centering recommendations for cross-sectional multilevel models from Enders & Tofighi (2007) and Enders (2013), in order to control for the relative within-cluster deviations. Multilevel analyses were conducted using the *lme4* and *nlme* packages in R (Bates et al., 2007; Pinheiro et al., 2017). I specified an autocorrelation structure of order 1 in order to allow residuals to be correlated over time, given the dependent nature of the multiple time-series data structure (Box et al., 2011). Unstandardized regression weights are reported. Significance testing of fixed effects was performed using *t* tests; the calculation of degrees of freedom and t-test for the multilevel models were estimated using a Satterthwaite approximation (Satorra & Bentler, 1994) in the package *lmerTest* (Kuznetsova et al., 2017).

Correlations between all study variables. Below is the correlation table for all key study variables. Interestingly, by looking at this table, I can probe the question: are people able to

accurately identify their motivations for self-disclosure in our study? It does seem that people are somewhat "successful" (defined as a significant, positive, moderate in strength relationship between two predictors) in accurately identifying their motivations. For example, there is a moderately strong correlation between sharing to feel good and share positivity, as well as sharing to connect and share responsiveness. This suggests that when people report that they are motivated to self-disclose to fulfil a certain goal (e.g., to feel good or to connect), the features of the share (e.g., positivity and responsiveness) reflect the initial motivation for sharing. However, for the motivations of sharing to manage impressions and sharing with no thought, I cannot draw the same conclusions, although I could predict that since people are moderately accurate in identifying the other two motivations, they are likely to be accurate in identifying these ones, too.

	Sharing to connect	Sharing to feel good	Sharing to manage impressions	Sharing with no thought	Share positivity	Share responsiveness
Sharing to connect	---	0.37*	0.43*	.02	.30*	.38*
Sharing to feel good	0.48*	---	.44*	.02	.24*	.29*
Sharing to manage impressions	0.21*	0.29*	---	.16*	.31*	.26*
Sharing with no thought	-0.03	-0.02	-0.07*	---	.15*	.10*
Share positivity	.27*	.24*	.20*	.17*	---	.37*
Share responsiveness	.28*	.23*	-.03	.06*	.34*	---

Table 2. Correlations between the four motivations for self-disclosure and characteristics of the disclosure. Above the diagonal is social media; Below the diagonal is face-to-face.
* $p < 0.05$.

Are there differences in affective well-being and social connection across communication contexts? I examined whether participants' reports of affective well-being and feelings of social connection were different during the time periods when they were reporting face-to-face versus social media self-disclosures. There were no significant differences between overall affect during the face-to-face reports (M = 5.98, SD = 2.23) versus social media reports (M = 6.04, SD = 2.21), t(12,992) -1.59, $p > 0.05$. There were also no significant differences in overall feelings of social connection (Face-to-face: M = 5.26, SD = 2.53; Social media: M = 5.31, SD = 2.54), t(13,020) -1.21, $p > 0.05$. Thus, the effect of self-disclosure motivation on affect and feelings of social connection is not due to variation in happiness and social connection between the communication context reporting weeks (e.g., it's not the case that people were simply happier during the face-to-face reporting week).

Are there differences in characteristics of self-disclosures in face-to-face versus social media contexts? To examine if perceived self-disclosure positivity and audience responsiveness was different for self-disclosures made face-to-face versus on social media, I performed a series of Welch's t-tests, assuming unequal variances because of the disparity in number of self-disclosures made face-to-face versus on social media.

Positivity was significantly higher for social media self-disclosures (M = 6.63, SD = 2.64) as compared to those made face-to-face (M = 6.17, SD = 2.44), t(801.53) = -3.90, $p < 0.001$, d = -0.18 [-0.26, -0.09]. Conversely, audience responsiveness was significantly higher for self-disclosures made face-to-face (M = 7.27, SD = 1.91) versus on social media (M = 6.13, SD = 2.57), t(723.70) = 10.29, $p < 0.0001$, d = 0.47 [0.38, 0.55]. These results suggest two key differences between communication contexts: users share more positive thoughts and feelings on social media, which aligns with past research showing that people are more likely to express

positive emotions on Facebook and negative emotions face-to-face (Qiu et al., 2012). However, people perceive others to be more responsive to what they share when it's done face-to-face.

How often do people endorse different motivations for self-disclosure in face-to-face versus social media contexts? I again performed a series of Welch's t-tests to examine if people rate the extent to which four possible motivations drove their most recent self-disclosure based on whether the disclosure was made face-to-face or on social media (Table 2).

People were significantly more likely to be motivated by a desire to connect with others when disclosing in face-to-face contexts as opposed to on social media, $t(763.93) = 6.50$, $p < .001$, $d = 0.29$ [0.20, 0.38]. Surprisingly, people were also significantly more likely to self-disclose without a motivation or with no thought in face-to-face contexts, despite work (e.g., Suler, 2004) hypothesizing that on social media, users have restricted self-awareness and may be more likely to share in a disinhibited manner, $t(784.27) = 4.77$, $p < .001$, $d = 0.22$ [0.13, 0.30]. In line with prior literature (e.g., Walther, 2007) about the strategic self-presentation affordances that social media provides, people were indeed significantly more likely to want to self-disclose in order to make a good impression on social media as opposed to face-to-face, $t(805.51) = -3.42$, $p < .001$, $d = -0.15$ [-0.24, -0.07]. People did not report being significantly more or less motivated to disclose in order to feel good between face-to-face and social media contexts, $p = .19$.

	Social media ***N*** **= 589**	**Face-to-face** ***N*** **= 2,942**	**t**	**df**	***p*-value**	**Cohen's *d***
To socially connect	*M* = 5.76 *SD* = 2.98	*M* = 6.61 *SD* = 2.51	6.50	763.93	< .001	0.29 [0.20, 0.38]
To make a good impression	*M* = 4.10 *SD* = 2.97	*M* = 3.64 *SD* = 2.77	-3.42	805.51	< .001	-0.15 [-0.24, -0.07]

To feel good	*M* = 6.07 *SD* = 2.77	*M* = 5.91 *SD* = 2.63	-1.30	*813.35*	.19	-0.06 [-0.15, 0.03]
With no thought	*M* = 5.88 *SD* = 3.10	*M* = 6.53 *SD* = 2.76	4.77	*784.27*	< .001	0.22 [0.13, 0.30]

Table 3. Welch's t-tests for self-disclosure motivations by communication context.

Does motivation for self-disclosure *face-to-face* predict changes in well-being over time? First, I looked at how one's motivation for self-disclosure in face-to-face contexts impacts affective well-being and feelings of social connection over time. Specifically, as with examining this question on social media, I asked whether T_2 happiness and feelings of social connection were predicted by $T_{1\text{-}2}$ motivation for self-expression face-to-face (simultaneously entered), controlling for T_1 happiness or feelings of social connection at level-1 of the model.

As indicated in Table 4 below, all motivations for self-disclosure (to connect, to make a good impression, to feel good, and with no thought) in face-to-face contexts significantly improved affective well-being and feelings of social connection over time.

Motivation	B	SE	t	df	*p*-value
To socially connect					
• Affective well-being	0.34	0.05	7.40	1882	< .001*
• Social connection	0.28	0.05	5.66	1892	< .001*
To make a good impression					
• Affective well-being	0.11	0.04	2.69	1882	.007*
• Social connection	0.13	0.05	2.90	1892	.003*
To feel good					

• Affective well-being	0.27	0.05	5.97	1882	< .001*
• Social connection	0.25	0.05	5.29	1892	< .001*
With no thought					
• Affective well-being	0.11	0.04	2.53	1882	.01*
• Social connection	0.13	0.04	2.89	1892	.004*

Table 4. Effects of motivations for self-disclosure face-to-face on subjective well-being and social connection over time.
* $p < .05$.

Does motivation for self-disclosure on *social media* predict changes in either affective well-being or feelings of social connection? Next, I examined how one's motivation to self-disclosure on social media impacts affective well-being and feelings of social connection over time. Specifically, I examined in two multilevel models (one for each dependent variable) whether T_2 affective well-being or feelings of social connection were predicted by $T_{1\text{-}2}$ motivation for self-disclosure on social media (simultaneously entered), controlling for T_1 happiness or feelings of social connection at level-1 of the model. As indicated in Table 3 below, only when participants reported being more motivated to self-disclose on social media in order to feel good did they show an increase in their affective well-being (B = 0.28, $p = .02$) and feelings of social connectedness (B = 0.31, $p = .03$) over time. For all other motivations, there was no significant effect on changes in affective well-being or feelings of social connectedness over time.

Interestingly, overall, participants who reported sharing once or more on social media since the last text, as compared to those who did not report sharing on social media, showed significant increases in affective well-being (B = 0.51, $p < .001$) and feelings of social connect (B = 0.51, $p < .001$) over time. However, due to the wording of the survey, I can't conclude

whether the comparison group is either people who didn't self-disclose, but still used social media (e.g., passive usage; Verduyn et al., 2015) or people who didn't self-disclose and also didn't use social media. Given how much people use social media in general, I would hypothesize that the comparison group is the former. Cautiously, what this might suggest is that it is worthwhile to share anything on social media, if you are using social media already, in order to increase well-being, regardless of your motivation. This would be an important area for further study.

Motivation	**B**	**SE**	**t**	**df**	***p*-value**
To socially connect					
• Affective well-being	0.11	0.09	1.22	233	.22
• Social connection	0.12	0.11	1.15	234	.25
To make a good impression					
• Affective well-being	0.13	0.11	1.19	233	.23
• Social connection	0.25	0.13	1.97	234	.05
To feel good					
• Affective well-being	0.28	.012	2.37	233	.02*
• Social connection	0.31	0.14	2.21	234	.03*
With no thought					
• Affective well-being	0.05	0.10	0.50	233	.62
• Social connection	0.23	0.11	1.99	234	.05

Table 5. Effects of motivations for self-disclosure on social media on subjective well-being and social connection over time.
* $p < 0.05$.

Does motivation for self-disclosure differ between *face-to-face* and *social media* conditions in predicting changes in well-being over time? Next, I examined how the effects of one's motivation for self-disclosure interacts with whether the self-disclosure took place face-to-face or on social media. Only the motivation of sharing to connect had a significant interaction with communication context in predicting social connection, $B = -.26$, $t(2303.64) = -2.47$, $p = 0.01$. Specifically, the effect of the motivation of sharing to connect was significantly stronger in the face-to-face context. All other motivations did not significantly differ by communication context.

Does self-disclosure *overall* predict changes in well-being over time? Finally, I examined, agnostic to the motivation for self-disclosure, how simply self-disclosing face-to-face or on social media influenced affective well-being and feelings of social connection over time. First, participants who reported sharing once or more face-to-face since the last text, as compared to those who did not report sharing face-to-face, showed significant increases in affective well-being ($B = 0.76$, $p < .001$) and feelings of social connection ($B = 1.34$, $p < .001$) over time.

Similarly, participants who reported sharing once or more on social media since the last text, as compared to those who did not report sharing on social media, showed significant increases in affective well-being ($B = 0.51$, $p < .001$) and feelings of social connection ($B = 0.51$, $p < .001$) over time. Notably, the magnitude of the effect of sharing is *much larger* in the face-to-face condition, particularly as it relates to feelings of social connection. Thus, the results suggest that any self-disclosure at all, agnostic to motivation for sharing, can improve well-being-related outcomes.

Part II: Applying dimensionality reduction for the four self-disclosure motivations

Correlations between self-disclosure motivations. When calculating the correlations between all key study variables (Table 2), I noticed that many of the self-disclosure motivations were significantly correlated with one another (see Table 6 below). Specifically, the motivations of *sharing to connect*, *sharing to feel good*, and *sharing to manage impressions* were positively and significantly correlated with one another in both the face-to-face and social media conditions. Upon review of the correlations, I hypothesized that the four motivations for self-disclosure might be able to be broken down into two components: component 1 would consist of the significantly positively correlated motivations of *sharing to connect*, *sharing to feel good*, and *sharing to manage impressions* and component 2 would consist of the motivation of *sharing with no thought*. In order to test this hypothesis, I conducted a principal component analysis of the four motivations for self-disclosure separately for the face-to-face and the social media reporting conditions.

	To socially connect	To feel good	To manage impressions	With no thought
To socially connect	---	0.37*	.43*	.02
To feel good	0.48*	---	.44*	.02
To manage impressions	0.21*	0.29*	---	.16*
With no thought	-0.03	-0.02	-0.07*	---

Table 6. Correlations between four motivations for self-disclosure. Above diagonal is correlations for social media observations; Below diagonal is for face-to-face observations.
* $p < 0.05$.

Principal component analysis. Given these significant correlations between the four motivations for self-disclosure, I explored if the dimensions of these motivations could be

reduced using principal component analysis (Abdi & Williams, 2010). Using principal component analysis, I extracted two components, which explained 62% (face-to-face) and 64% (social media) of the variation among the motivation predictors. I rotated[2] loadings and found evidence for extracting the two named components: *deliberate* and *impulsive* self-disclosure (Table 7).

I named these components deliberate and impulsive self-disclosure based on past literature about dual-process frameworks. As mentioned in the introduction, dual-process frameworks suggest that human thought, feeling, and behavior are the product of an interplay between two systems: an impulsively driven, reflexive "hot" system that consumes minimal effort and a deliberatively driven, reflective "cool" system that requires effort to operate (Chaiken & Trope, 1999). Considering the two components suggested by the principal component analysis, it seems that component one maps onto the deliberate, "cool" system and component two maps onto the impulsive, "hot" system. Component one is comprised of three motivations that all have thoughtful and deliberate goals – to socially connect, to feel good, and to manage impressions. On the other hand, component two is comprised of the one motivation that does not have a clear, thoughtful goal; indeed, the motivation is defined as self-disclosing "with no thought."

	Deliberate Self-Disclosure	Impulsive Self-Disclosure
To socially connect	0.77 0.79	
To feel good	0.78 0.83	

[2] I used oblique (*oblimin*) rotation in order to maximize variables loadings onto one component while minimizing its loadings onto all other components.

To manage impressions	0.58	0.78
With no thought		.98 0.99

Table 7. Principal component analyses loadings. Above diagonal is loadings for social media observations; Below diagonal is for face-to-face observations.

Do the dual-process motivations (i.e., deliberate and impulsive) for self-disclosure in *face-to-face* contexts predict changes in well-being over time? I examined in two multilevel models (one for each dependent variable) whether T_2 affective well-being or feelings of social connection were predicted by $T_{1\text{-}2}$ motivation (deliberate or impulsive) for face-to-face self-disclosure (simultaneously entered), controlling for T_1 happiness or feelings of social connection at level-1 of the model.

Both the deliberate motivation and the impulsive motivation for self-disclosure significantly improved affective well-being and feelings of social connection over time (Table 8). These results are aligned with the findings of conducting this analysis inputting the four motivations separately (see Table 4).

Motivation	**B**	**SE**	**t**	**df**	***p*-value**
Deliberate self-disclosure					
• Affective well-being	0.47	0.03	11.38	1884	< .001*
• Social connection	0.42	0.04	9.50	1894	< .001*
Impulsive self-disclosure					
• Affective well-being	0.11	0.04	2.69	1884	.007*
• Social connection	0.13	0.05	2.97	1894	.003*

Table 8. Effects of the dual-process motivations for self-disclosure **face-to-face** on subjective well-being and social connection over time.
* $p < 0.05$

Do the dual-process motivations (i.e., deliberate and impulsive) for self-disclosure on *social media* predict changes in well-being over time? Again, I examined in two multilevel models (one for each dependent variable) whether T_2 affective well-being or feelings of social connection were predicted by $T_{1\text{-}2}$ motivation (deliberate or impulsive) for self-disclosure on social media (simultaneously entered), controlling for T_1 happiness or feelings of social connection at level-1 of the model.

Only the deliberate motivation significantly improved affective well-being and feelings of social connection over time; the impulsive motivation had no significant effect on either dependent variable (Table 9). Notably, when assessing the motivations separately, only the motivation of sharing *to feel good* was a significant predictor, but with slightly smaller effects (B = 0.28 for affective well-being and B = 0.31 for social connection) as compared to the component of deliberate sharing (B = 0.41 for affective well-being and B = 0.45 for social connection). This suggests that there is value in collapsing the more thoughtful motivations into one component, in terms of an increased effect on the outcomes.

Motivation	**B**	**SE**	**t**	**df**	***p*-value**
Deliberate self-disclosure					
• Affective well-being	0.41	0.11	3.67	212	< .001*
• Social connection	0.45	0.13	3.37	213	< .001*
Impulsive self-disclosure					
• Affective well-being	0.06	0.10	0.54	212	.59
• Social connection	0.13	0.12	1.10	213	.27

Table 9. Effects of the dual-process motivations for self-disclosure on **social media** on subjective well-being and social connection over time.
* $p < 0.05$

Part III: Moderation by individual difference variables and self-disclosure characteristics

The main effects of motivation for sharing on well-being in both face-to-face and social media contexts were not moderated by individual differences as measured by the baseline questionnaires (extraversion, self-esteem, need to belong, and self-control). Additionally, supplemental analyses did not find, both for face-to-face and social media shares, that self-reported positivity of the self-disclosure moderated any outcome. However, perceived audience responsiveness did significantly moderate the effects in the face-to-face condition of being motivated to share to feel good on affective well-being, $B = 0.12$, $SE = 0.05$, $t(1955) = 2.56$, $p = .01$, and on social connection $B = 0.15$, $SE = 0.05$, $t(1947.98) = 2.95$, $p = .003$, such that the perception of higher audience responsiveness led to stronger effects on the two outcomes.

Although only marginally significant, at $p = .06$, I want to draw attention to the finding that the responsiveness of one's audience on social media somewhat moderated the effects of sharing to connect, $B = 0.15$, $SE = 0.08$, $t(318.9) = 1.87$, and sharing with no thought, $B = 0.15$, $SE = 0.08$, $t(322.6) = 1.92$, on affective well-being. Specifically, these marginal results suggest that increased audience responsiveness may increase the magnitude of the effects of sharing for these two motivations on happiness (see Figures 2 and 3 below). This suggests that perhaps having a supportive audience on social media can strengthen the effects of self-disclosure on well-being.

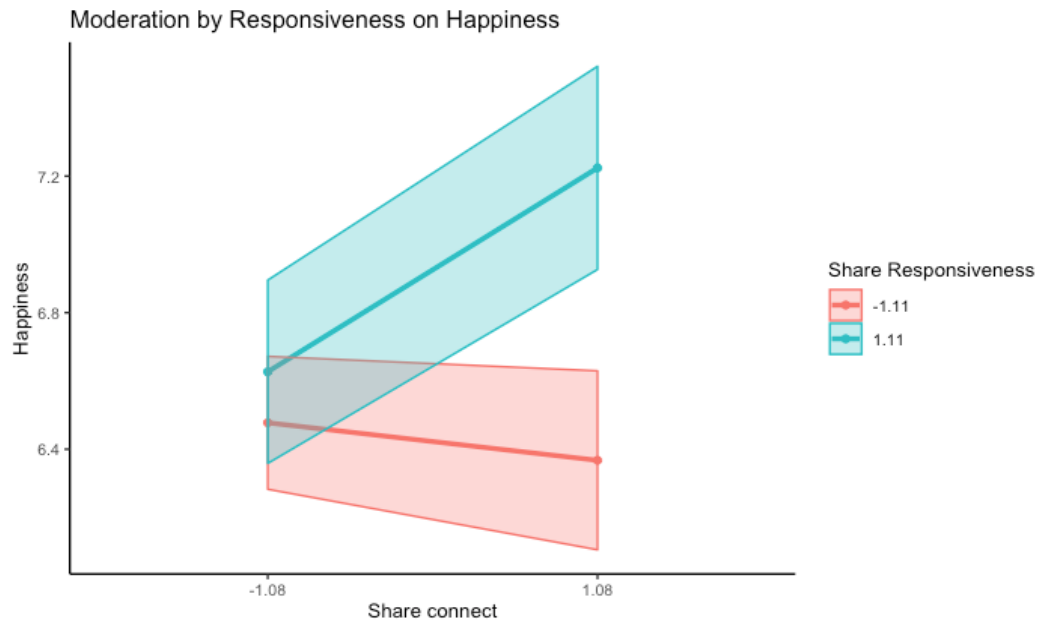

Figure 2. Moderation by audience responsiveness on the outcome of affective well-being ("happiness") for the motivation of sharing to connect in the social media condition.

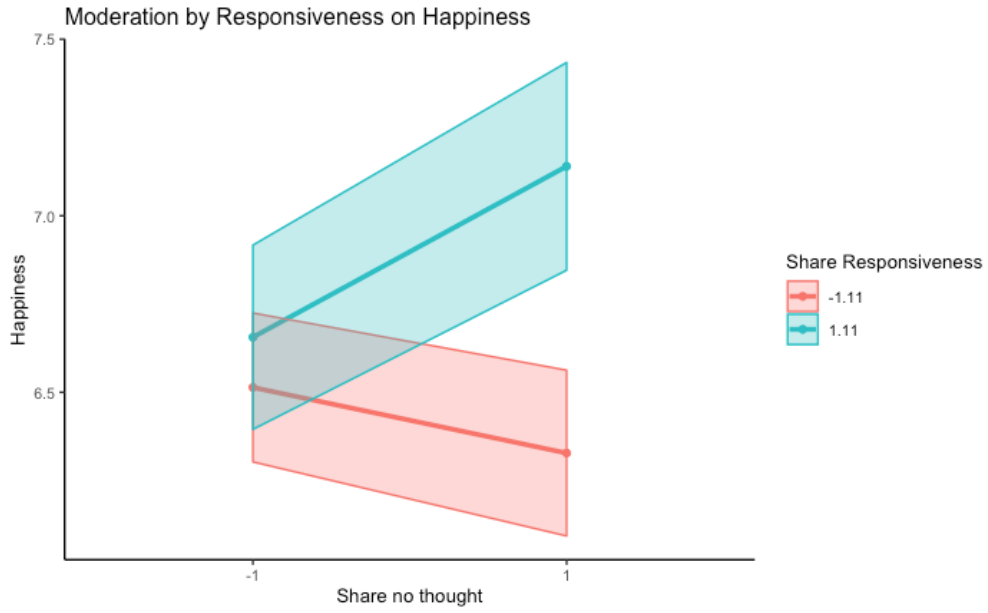

Figure 3. Moderation by audience responsiveness on the outcome of affective well-being ("happiness") for the motivation of sharing with no thought in the social media condition.

Additionally, differences between communication context were only significant when predicting affective well-being from motivation to socially connect. What this result reflects is that the standard errors for the predictors in the social media context were often double the size of the standard errors for the predictors in the face-to-face context. Although I collected four individual difference measurements as potential moderators – extraversion, self-esteem, need to belong, and self-control—these individual differences did not significantly moderate results. Nonetheless, the larger standard errors in the social media condition reflect larger variability that could be due to individual difference measures that I did not include in the present study. Future research is needed to explore different individual difference possibilities.

Discussion

The present research examined if self-disclosures made face-to-face versus on social media differentially influence feelings of affective well-being and social connection over time. Given the many benefits of self-disclosure in face-to-face contexts (e.g., Collins & Miller, 1994, Laurenceau et al., 1998), the possibility that disclosures made on social media are similarly beneficial is intriguing. The present research confirmed that face-to-face self-disclosures are advantageous—no matter one's motivation for sharing, face-to-face disclosures significantly increased affective well-being and feelings of social connectedness over time. Although the magnitude of these effects varied—with the motivation of sharing to make a good impression and sharing with no thought having smaller effects—overall, face-to-face self-disclosures significantly increased well-being-related outcomes over time.

The same did not hold true on social media. Instead, I found that only when one reports a motivation to self-disclose in order to feel good did sharing significantly increase affective well-being and feelings of social connectedness. Not even the motivation of disclosing to socially

connect on social media predicted increases in feelings of social connectedness. However, when conducting additional analyses modeling the four motivations as two principal components—deliberate and impulsive—endorsing a deliberate motivation for sharing did significantly predict increases in affective well-being and feelings of social connectedness over time. However, impulsive sharing had no significant effect. As a note, my findings did not suggest that self-disclosure on social media significantly *worsens* well-being over time. Other researchers have found evidence that sharing on social media harms mental health (Shakya & Christakis, 2017) and that passively consuming social media content without posting significantly decreases affective well-being (Verduyn et al., 2015). However, given the positive effects of face-to-face self-disclosures, only sharing on social media may not be a sufficient alternative.

Taken together, the implications of these results suggest that social media interaction is not a precise surrogate for offline interaction. This is an important finding given that Facebook states that its mission is to connect people and that the platform's design encourages public self-disclosure. However, the data do suggest that self-disclosure on social media can be useful, as long as one is strategic in their motivation for doing so, as well as have a clear motivation in mind to feel good by self-disclosing.

Limitations and future studies. First, due to a technical error during the data collection phase of the experience sampling study, our team was not able to collect the actual written Facebook and Twitter posts from our participants, as well as the number of likes and comments on each post. We did create this study with the intention to collect the actual content of what people posted online. But due to a coding error in Python, this social media data was not accurately recorded and therefore could not be analyzed. With this additional data, I would have assessed how accurately participants reported characteristics of their social media self-disclosure,

such as others' responsiveness (via likes and comments) and post positivity (via sentiment analysis). As a side note, we did not attempt to collect the equivalent data (i.e., what people say during self-disclosures) in the face-to-face condition, which would have allowed us the opportunity to make direct comparisons between the two. It would be hard, but not impossible, to collect this same data using an experience sampling method face-to-face (using audio-recording devices is one possibility, but the data would be laborious to code and analyze without automated systems).

In any case, insights from the actual social media posts—and how they compare to people's perceptions—would be interesting in light of the evidence suggested by the correlational data that people are accurately able to identify their motivations for sharing. With this additional social media post data, we could see if participants are also accurately able to identify the characteristics of their share. However, I would hypothesize this may be more difficult for our participants, as the latter requires making accurate predictions about how other people behave. Research consistently suggests that people commit many biases in judging how other people will react to them (e.g., Bohns, 2016; Bruk et al., 2018, just in online spaces). Past research suggests that sometimes people do not accurately perceive the amount of support they receive to their social media posts if they are suffering from depression (J. Park et al., 2016). I would like in future studies to examine if the motivation one has for sharing interacts with perceptions of support and actual support (via likes and comments) to predict changes in well-being over time. I would hypothesize, for example, that being motivated to share to connect on social media and then not receiving actual support could be significantly *worsen* feelings of social connection over time (as compared to not posting at all). However, this hypothesis is

based on the assumption that perceived support and actual support are significantly correlated, which in itself is a testable hypothesis for future studies.

An additional shortcoming of the present research is that I limited reporting of social media interactions to *public* Facebook and Twitter posts. These social media sites offer opportunities for private, more directed communications and messaging as well. Perhaps the public/private dimension of a social media self-disclosure influences the consequences of sharing on well-being and social connection over time. In future studies, to overcome this limitation, I could limit the definition of self-disclosure for our participants as private, directed communications (e.g., direct messages on Facebook and Twitter). Another option is that I could collect both kinds of self-disclosures, those made publicly and privately, and then compare the two. One limitation of this method is that people may not be willing to share the content of their private messages (whereas we believed for the present study that participants would be more willing since the posts we intended to collect were already public).

In these future studies, I hypothesize that private sharing on social media leads to more beneficial outcomes as it more closely matches the process of sharing face-to-face (which in the current research was always beneficial). However, I would still expect the magnitude of these effects to be smaller than those for face-to-face disclosures, based on the data included in this study. I would predict that some features of social media, such as asynchronous communication and lack of nonverbal cues, would dampen some of these beneficial effects, whereas this would not be the case during private face-to-face self-disclosures.

Another limitation of the present research is that I did not ask participants to rate their feelings of closeness to the target(s) of their self-disclosure. It seems reasonable to imagine that face-to-face disclosures might have a higher probability of being made to a closer friend, while

social media disclosures might have a more diverse (and perhaps more socially distant) audience. Although research suggests that even small face-to-face interactions with strangers increase positive affect (Sandstrom & Dunn, 2014), I would predict that the benefits of face-to-face self-disclosure are even stronger when made to a close friend, and this effect may be similar on social media. Relatedly, past research by Burke and Kraut suggests that the effect of Facebook use on well-being depends on if one is sharing to "strong" or "weak" ties, which also informs this prediction (Burke & Kraut, 2016). In the next chapter, although in a different realm of interpersonal behavior, I will examine how relationship closeness influences how people communicate face-to-face versus on social media.

Finally, one of the largest limitations of the present research is perhaps in how I conceptualized motivation for self-disclosure in the beginning. I drew upon a body of prior research to propose a four-motivation framework for why people share, and I chose the four motivations because they appeared consistently in both face-to-face and social media literatures. These past studies often drew upon surveys and qualitative interviews (e.g., Bazarova & Choi, 2014; Nadkarni & Hoffman, 2012) to identify people's motivations for self-disclosure on social media.

However, first, I would propose that in future studies, our research team would conduct our own lab-based experiments and ask participants their motivations for sharing. Specifically, we could improve upon these prior studies by asking participants to report their motivation for self-disclosure *in-the-moment* of sharing. While these past studies asked people to reflect on past experiences, I think having participants self-disclose either face-to-face or on social media and then *immediately* reflect on why would help more accurately define what people's motivations truly are for self-disclosure.

Second, another approach I could have taken to conceptualize motivation for self-disclosure is to utilize the interpersonal theory of emotion regulation as a framework for understanding self-disclosure as a broader emotion regulation strategy altogether (Zaki & Williams, 2013). Specifically, this framework proposes that sharing our thoughts and feelings through self-disclosure to others allows us to effectively regulate our inner emotional states. I could extent this framework to the social media realm to predict that self-disclosure online is a way that we attempt to regulate our emotions with the support of other people. I say "attempt" because interpersonal emotion regulation may only be successful in face-to-face contexts. I would hypothesize that self-disclosure on social media *could* be an effective interpersonal emotion regulation strategy, but only as long as people can accurately perceive social support online.

Chapter 2: Relationship Closeness Impacts Hostility During Confrontations about Moralized Political Topics on Social Media

Introduction

In Chapter 1, I examined how people share their thoughts and feelings in face-to-face and social media contexts, finding that not only does the context of the self-disclosure matter, but also one's motivation for sharing. In this chapter, as in the previous chapter, I will examine how context may influence how people communicate, specifically when people share their views about moralized political topics. However, in this chapter, I turn from looking at the individual discloser to examining the effects of the *relationship between communication partners* sharing their thoughts and feelings. Specifically, I will take a close look at how a dyad's relationship closeness (as operationalized by network distance on Twitter) influences hostility (as operationalized by a popular toxicity algorithm) in direct tweets about moralized political topics.

Background

Imagine debating with your best friend the pros and cons of a common polarizing and moralized political topic: keeping abortion legal. Now, imagine debating the same topic with your dentist. Would these conversations feel the same? Would you say the same things to your best friend and your dentist? Likely, these two conversations would be and feel very different—in your ease of sharing your thoughts, how connected you felt during the conversation, and maybe even your willingness to voice certain opinions. You might even avoid engaging in such a discussion with your dentist, given the choice. It's evident, from this thought experiment and from multiple psychological studies (e.g., DePaulo & Kashy, 1998; Planalp, 1993; Planalp &

Benson, 1992; Savitsky et al., 2011; Tedeschi, 1986; Tice et al., 1995), that the relationship closeness between communication partners influences how the dyad thinks, feels, and behaves during their face-to-face interaction.

It's becoming evident that relationship closeness impacts the nature of social media interactions, as well. For example, a large body of work by Burke & Kraut suggests that it's critical to take into account *who* a user is communicating with, in terms of perception of relationship closeness, when understanding how social media communication may impacts a user's affective well-being over time (Burke et al., 2011; Burke & Kraut, 2013, 2016, 2014). Specifically, these researchers found that receiving targeted, personalized social media messages from close friends (i.e., "strong ties") is associated with significant improvement in well-being after using social media, while receiving messages from more distant acquaintances or strangers (i.e., "weak ties") is not. While this work highlights the importance of understanding relationship closeness in order to understand the affective consequences of social media use, this work has not yet been extended to looking at how relationship closeness impacts how people *actually communicate* on social media. The aim of the present research is to fill this gap by considering how people converse, specifically about moralized political issues, on social media to friends versus acquaintances.

The politics of discussing politics

Conversations, whether they take place on social media or face-to-face, can span a range of topics. Why focus on how people discuss controversial political topics? Research from multiple fields including psychology, political science, and communications suggest that disagreements in particular may look different in offline versus online spaces.

On the one hand, studies focusing on offline communications have consistently found that people often attempt to avoid face-to-face confrontations and other negative social interactions, both toward acquaintances and strangers (Fisher, 1979; Milgram & Sabini, 1978; Rosen & Tesser, 1970; Sabini et al., 2000) and friends (Argyle & Henderson, 1984). Potential reasons for this effect are that we are motivated by, among other drives, strong norms to conform (Asch, 1951), cooperate (Fehr & Fischbacher, 2004), and build social connections with others (Baumeister & Leary, 1995; Rusbult & Buunk, 1993). Another possible explanation is that cues to another person's mental state are more salient during verbal face-to-face communication (e.g., paralinguistic cues such as intonation and emphasis; nonverbal cues such as facial expressions and body language), and these cues, when present, foster more empathy, less dehumanization, and less stereotyping between conversation partners (Epley & Kruger, 2005; Kumar & Epley, 2020; Schroeder & Epley, 2015). For example, research has shown that hearing a stranger's voice (as opposed to seeing his or her words in text) leads to less dehumanization during disagreements about polarizing political topics, such as abortion (Schroeder et al., 2017). Cues that facilitate empathy seem to be less salient (or at the least more ambiguous) during text-based messaging, a common form of communication online. However, communicators may not perceive this gap; research shows that people are overconfident in their ability to communicate empathic cues such as tone online (Kruger et al., 2005) In sum, expressing disagreement, especially disagreement characterized by hostility, is fairly *un*common during face-to-face conversations (Mutz, 2006). Instead, people often choose to avoid these conversations or stay silent as opposed to speaking up with contradictory opinions (Eliasoph, 1998; Noelle-Neumann, 1974).

On the other hand, a growing body of research suggests that people are more vocal and hostile when it comes to speaking up *online*, especially about polarizing political beliefs (for a recent review about the Internet's broad effects on political behavior, see Zhuravskaya et al., 2020). While not all research has found the Internet to have a negative influence—for example, one study found that discussing politics on social media can increase political engagement offline (Lane et al., 2017)—other research has raised cause for concern about the way people interact online when discussing politics. For example, studies have shown that discussing political topics in online spaces leans toward being uncivil and aggressive (Barnidge, 2017; Coe et al., 2014). To understand why, we can look to a large body of literature that indicates online interactions, more generally defined, are often more hostile (Cheng et al., 2017; Englander & Muldowney, 2007; Kowalski et al., 2014; Moore et al., 2012; Tokunaga, 2010; Whittaker & Kowalski, 2015) or more likely to incite moral outrage (Brady & Crockett, 2019; Crockett, 2017; Hofmann et al., 2014) than offline interactions. Furthermore, hostile online environments often beget more hostility. For example, research has shown users will conform to aggressive norms in blog commenting sections (Rösner & Krämer, 2016). Alarmingly too, hate speech on social media (such as anti-Semitism and anti-Muslim sentiment) has been linked to increases in *offline* violent crime against these groups (Müller & Schwarz, 2018, 2019; Williams et al., 2019).

But why might online interactions be more toxic than offline ones? First, as mentioned earlier, paralinguistic and nonverbal behavioral cues conveyed during face-to-face communication foster empathy between conversation partners. However, these cues are often ambiguous or absent during virtual text-based communication. A lack of these cues makes dehumanization more likely, which can then beget increased hostility during a disagreement (Haslam, 2006; Schroeder et al., 2017). Furthermore, not only are people less sensitive toward

other people's mental states during online communication, but they may also be less self-aware of the implications of their *own* disclosures, which can lead to more disinhibited disclosures to strangers (Joinson, 2001). This decrease in empathic verbal cues and public self-awareness, combined with a decrease in private self-awareness, can give rise to increased disinhibition when sharing one's thoughts and feelings on social media, as compared to face-to-face (Clark-Gordon et al., 2019; Joinson, 2001; Suler, 2004).

Despite its potential harms, disinhibition during online conversations is not exclusively negative. Disinhibition can lead to a higher quantity of self-disclosure and, as I discussed in chapter 1, increased self-disclosure can help strengthen relationships and improve well-being (e.g., Collins & Miller, 1994; Cozby, 1973; Laurenceau et al., 1998). For example, researchers have found online disinhibition prompts more disclosures about sensitive topics like mental illness or abuse, and these increased disclosures lead to more audience engagement and support (Andalibi et al., 2018; Balani & De Choudhury, 2015; De Choudhury & De, 2014). However, online disinhibition during more confrontational interactions, such as debating a moral dilemma with an anonymous stranger, has be shown to increase communication toxicity (i.e., more hostile statements, name-calling, swearing, insults, and threats) (Lapidot-Lefler & Barak, 2012).

This prior research helps elucidate why conversations, in particular about polarizing topics like politics, may be more confrontational online. Yet none of the previous research has measured relationship closeness between conversation partners as a predictor of this hostile behavior. In fact, much of the research on online disinhibition has focused exclusively on communication between anonymous strangers (e.g., Clark-Gordon et al., 2019; Joinson, 2001).

I predict that one's relationship with a conversational partner will likely impact the toxicity of their conversation, given past research that shows we converse differentially

depending on whether we are speaking to a friend or acquaintance (e.g., Planalp, 1993; Planalp & Benson, 1992). Specifically, I expect that close friends will be less hostile during conversations. Furthermore, the above research has not directly pitted how people confront close others or acquaintances—best friends or dentists—in online *versus* offline spaces (much of the research looks at how people behave *within* offline or online contexts, but not both). Filling in these gaps in the research is critical given that billions of people use social media to share and discuss ideas, and that one in five Americans use social media as their primary source for political news (Mitchell et al., 2020).

How we respond to moral transgressions and the importance of relationship closeness

Offline and online, people are strongly motivated to maintain relationships with close others (Burke & Kraut, 2014; Canary & Stafford, 1994; Ellison et al., 2014). We see close others as overlapping with our own sense of self (Aron et al., 1991). Close interpersonal relationships in turn have norms governing the inhibition of uncooperative or harmful behavior (Argyle & Henderson, 1984).

However, when people encounter moral violations, they are motivated to express their outrage by responding to the transgressor (Schein & Gray, 2016, 2018). This is due to a higher-order, evolutionary-based, motive to enforce the norms of one's in-group, even if it is costly to the individual (Fehr et al., 2002; Fehr & Fischbacher, 2004; Henrich & Boyd, 2001). Online, people are more likely to come across moral outrage-inducing material than in-person (Hofmann et al., 2014). This more frequent exposure online may increase people's likelihood to express their outrage to another person in this context. Indeed, research suggests that responding to moral transgressions on social media tends to be easier and more frequent (Brady & Crockett, 2019).[3]

[3] Paradoxically, viral moral outrage on social media actually elicits more sympathy for the transgressor (Sawaoka & Monin, 2018).

One type of moral outrage-inducing material people may come across on social media is posts and messages related to polarized political beliefs. In recent decades, politics has become an increasingly moralized issue, a phenomenon termed "political sectarianism" (Finkel et al., 2020). Based on this, I would predict that if one confronts another user who holds opposing political beliefs, which engenders strong moralized identification, on social media – a context where already more moral outrage-inducing content exists and also in which responding in these situations is easier (Brady & Crockett, 2019; Crockett, 2017) – these factors compound and influence *even more* hostility during this confrontation on social media than in-person.

But let's take a step back and ask: what happens if this person with a differing political opinion is a close friend? Foundational work in psychology suggests that people value the welfare of their friends and view the act of protecting close others as an extension of protecting the self (Aron et al., 1991; Trivers, 1971). So, for example, when a friend commits an unethical behavior, people are less likely to report this action to a third party (Waytz et al., 2013). If someone behaves inconsiderately during an interaction, people make more accommodations to forgive a friend than an acquaintance (Rusbult & Buunk, 1993). Indeed, this tendency extends to punishment, with people being more lenient in punishing moral transgressors if those transgressors are friends (Earp et al., 2020, p. 2020; Weidman et al., 2019). Based on this past body of research, I would predict that relationship closeness overrides one's motivation to respond in a hostile manner to another person expressing an antagonistic belief about a moralized political topic—if the person expressing that belief is a friend. And I predict this tendency would persist on social media.

Testing initial hypotheses with prior experimental research

In a set of six experimental studies (N = 1,449) for my master's thesis (Chandhok et al., in prep), I systematically examined the interplay between communication context, relationship closeness, and expectations about how one would think, feel, and behave during a hypothetical political confrontation about abortion or climate change. For this series of studies, I defined **confrontation**, informed by past related research (Czopp et al., 2006; Czopp & Monteith, 2003), as the act of speaking up about a disagreement one has with another person, particularly about a sensitive or controversial topic.

For this set of experiments, I followed a general research design (see Figure 1), which I then developed and individualized for each study based on that specific study's aim. I will describe the general research design here. Using Qualtrics to host the survey, participants first were randomly assigned to envision and name (1) someone who is very *close* to you, i.e., a friend, or (2) someone who is a very *distant* acquaintance. I asked participants to write the name of the person they were thinking of and their relationship to them (e.g., "my best friend Shawn" for the close condition or "my pharmacist Gus" for the distant condition).

Next, participants were instructed to read the following confrontational scenario:

> When reading the news, you come across an opinion article written by a local writer about climate change *(abortion)*. Not only do you disagree with the writer's point of view about climate change *(abortion),* but the writer has also drawn upon flimsy and false evidence to support their claims.

After reading this scenario, participants were randomly assigned to (1) imagine telling the writer *in-person* or *on the phone*[4] why they disagreed with the article or (2) imagine writing a *social media* post on Facebook about why they disagreed with the article. Rather than give

[4] The pattern of results was the same whether participants were in the in-person or phone call condition.

participants the option of choosing whether or not to confront the writer, we asked participants to imagine that they in fact had taken this confrontational action. This decision was based on prior research about people's preference to avoid confrontation when possible (Fisher, 1979; Rosen & Tesser, 1970) and a pilot study of this manipulation indicating people would choose to be *hypothetically* non-confrontational if given the option (which could also be attributable to a favorable response bias in participants as well).

Finally, participants were asked to self-report their comfort (i.e., "Describe how you would feel in this scenario." 1 – *Extremely uncomfortable*, 7 – *Extremely comfortable*) and perception of exerted effort (i.e., "Describe the effort it would require to take this action" (1 – *Extremely effortless*, 7—*Extremely effortful*) with the confrontational scenario. Additionally, participants reported the likelihood that they would take this confrontational action ("How likely would you be to act in the manner described in the hypothetical scenario?" 1 – *Extremely unlikely*, 7 – *Extremely likely*).

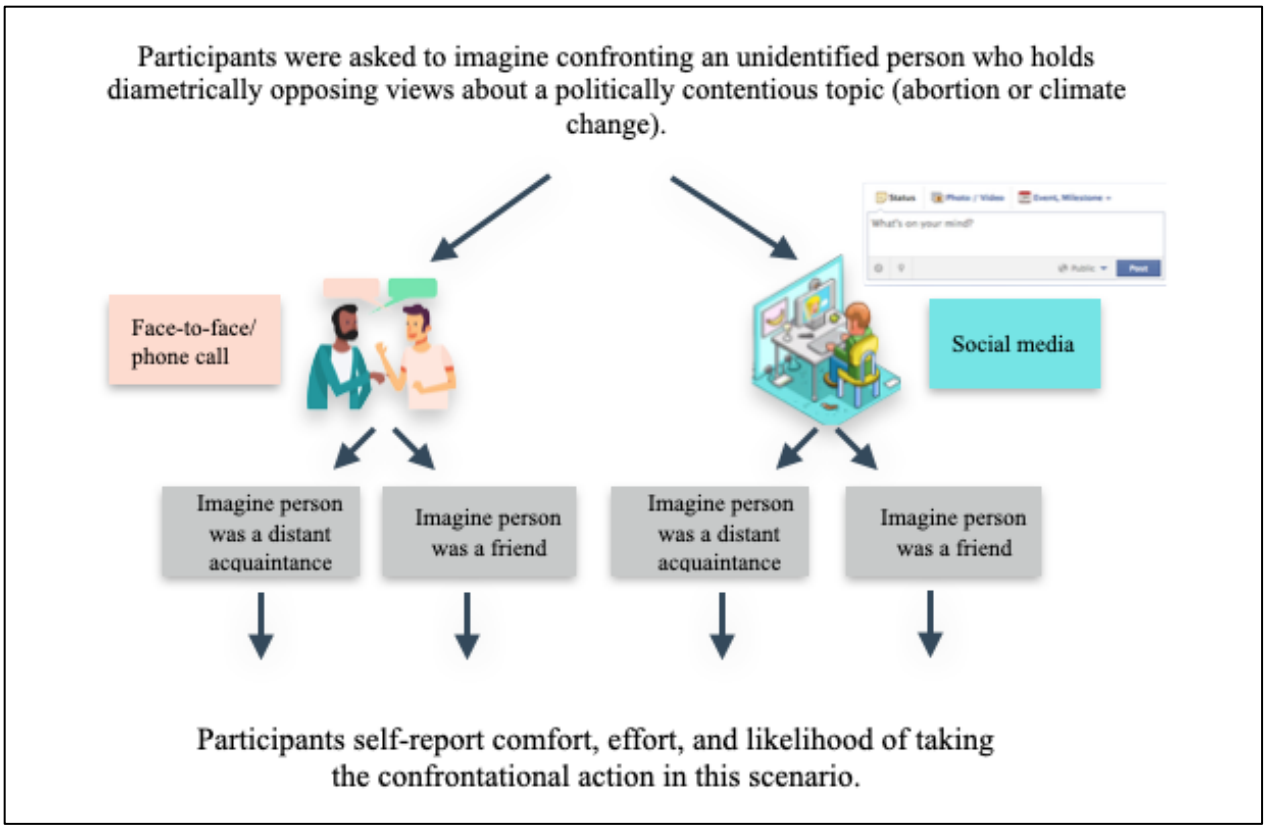

Figure 4. General outline of research design for all studies included in master's thesis.

In a meta-analysis of the effect sizes across the six studies, I found that relationship closeness significantly influences people's feelings and decisions about how to behave when confronting others about moralized political topics in offline and social media contexts. Specifically, collapsing across all six studies, people are significantly more likely to confront a distant acquaintance on social media (d = 0.28), finding this action significantly less effortful (d = -0.87) and more comfortable (d = 0.40). However, this effect reverses when people consider confronting a close friend. Participants found social media confrontation for someone close to be significantly more effortful (d = 1.02) and less comfortable (d = -0.86) and were less likely to do so (d = -0.93).

Extending experimental research to real-world social media interactions

This initial set of six studies provides consistent evidence that increased confrontational behavior on social media may be a function of relationship distance, such that people are more likely to be confrontational to distant acquaintances on social media. For the present study, I sought to replicate and extend these findings in the real-world by looking at actual Twitter conversations about confrontational topics. Moving this examination to Twitter has a number of benefits.

First, one of the limitations in the set of experimental studies is that I asked participants to consider hypothetical scenarios rather than truly engage in such a conversation. Indeed, an early speedbump I ran into, as previously mentioned, was that in pilot testing, participants would *hypothetically* choose *not* to participate in a confrontational scenario, given the option. How then could we translate these hypothetical findings to the real-world? Asking participants in a more ecologically valid, behavioral in-lab experiment to confront another person seemed risky, in terms of us not being able to observe enough incidences of confrontation, as our hypothetical data suggested. The benefit of using Twitter data is that these conversations had already naturally occurred. Furthermore, in this real-world, non-experimental setting, I could examine the ecological validity of my previous experimental findings, as well as have access to a large sample of confrontational conversations, on a spectrum of hostility. However, a drawback of moving this inquiry to Twitter is that I am only looking at one branch of these experimental findings – those confrontations that occur in the social media context.

Relatedly, another limitation of the experimental studies was that I did not directly measure hostile behavior during the confrontational scenario (since participants were asked to think hypothetically about their expectations rather than truly respond). I only measured expectations and perceptions about how difficult it *might* be to engage in the confrontation. With

the Twitter data, I was able to operationalize and measure hostility based on a toxicity algorithm applied to the language of the tweets.

I also want to highlight that confrontation, although categorized by disagreement, can have a range of associated hostility (which is defined as a number of related behaviors such as being disrespectful, insulting, belittling, or threatening to another person) (Lapidot-Lefler & Barak, 2012). For example, in presidential and Congressional debates over the years, we can see both civil and uncivil confrontations. Indeed, one of my hopes for the real-world application of this research is informing how people can have more civil and respectful confrontations.

In the next section, I present a confirmatory, crowdsourcing-based study examining how people engage in direct communication with another person on Twitter moralized, political topics. Specifically, I examine how relationship closeness, measured by two algorithms applied to Twitter, affects the hostility of these conversations. Specifically, I operationalize hostile behavior as toxicity in the language of the tweets, as measured by a machine learning algorithm explained in more detail in the methods section. I predict that the larger the social distance between two Twitter users publicly discussing controversial topics, the more likely they will be to use toxic language in the tweet.

Method

Design. I collected 283,587 tweets from October 1, 2017 to February 29, 2020 using the University of Michigan's Twitter Decahose dataset.[5] The tweets were collected based on the inclusion criteria of (1) being a directed tweet, (2) being a tweet containing phrases about abortion or climate change, (3) *not* being a tweet from a verified user. I will explain these inclusion criteria in more detail below.

[5] The Decahose dataset is a 10% random sample archive of all public tweets; see https://midas.umich.edu/twitter-decahose-data/ for more information.

First, I only included directed tweets that referenced a single other Twitter user's handle. Directed tweets are tweets that include the "@" symbol, which is the symbol for referencing another user's handle (i.e., their username). Non-directed tweets are aimed at one's broad Twitter audience rather than one specific user. For example, "Happy holidays!" is a non-directed tweet, but "Happy holidays @susannahs_dad!" is because it includes a direct reference to the user represented by the handle, "@susannahs_dad." I included only directed tweets because these tweets more closely reflect one-on-one communication, the type of behavior I examined in the prior set of experimental studies.

Second, I only included tweets that used one or more *phrases*, but not hashtags, about particular topic areas (which I will expand upon below), in line with prior recommendations about using Twitter for psychological research (see Appendix for the phrases used in the data collection and Table 1 for the broad categories of types of tweets collected) (Murphy, 2017). I considered collecting tweets based on the inclusion criterion of having hashtags (for example, tweets that included #abortion or #climatechange). Hashtags are commonly used by psychological researchers to collect a corpus of tweets about specific topics (recent examples include Bogen et al., 2019; García-Ramírez et al., 2019) However, a review of past literature suggested that hashtags have additional layers of meaning beyond mere topic identification; these additional functions include influencing, signaling, reaching a wide audience, and even sarcasm (Erz et al., 2018; Rauschnabel et al., 2019; Scott, 2015; Zappavigna, 2015). I decided to only use tweets based on the inclusion of phrases (e.g., "abortion") and excluding tweets that included any hashtags because non-hashtagged tweets more closely align with one-on-one communication, again the type of behavior I examined in the previous experimental studies.

Finally, I did not include tweets from or directed at "verified" Twitter users, which is a Twitter badge to denote that a user account of public interest is authentic (e.g., @katyperry is verified as being Katy Perry's authentic Twitter account).[6] Verified accounts are usually reserved for celebrities or businesses, so removing these types of users helped increase confidence that the tweets in the present study were both to and from the more average Twitter user.

After excluding tweets based on the above criteria, I was left with 213,304 tweets in total for subsequent analyses.

Tweet topics. Initially, I chose the two topics I examined in earlier studies (abortion and climate change), but then I expanded it to 19 topics in order to examine a range of controversial topics (i.e., topics that are likely to foster disagreement or confrontation), not only those that are political and moralized. To choose a sample of controversial topics beyond abortion and climate change, I turned to past academic research involving tweet collection on controversial topics (e.g., Barberá, 2015; Barberá et al., 2015), topics included in recent surveys from the Pew Research Center to measure views important to American voters[7], and a qualitative examination of popular trending topics on Twitter based on web research.

I collected tweets about a range of controversial issues in three broad categories, as well as an additional control category, for a total of four categories (see Table 1). First, I collected tweets that are **political and moralized**, as in the first iteration of tweet collection (i.e., abortion and climate change), but I also expanded data collection to include new topics such as the Black Lives Matter movement and gun control. The idea behind doing this was to increase power in the subsequent analyses, as well as control for any idiosyncrasies that may be due to the specific topic.

[6] https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts
[7] https://www.pewresearch.org/topics/

Second, I collected tweets that were **non-political but moralized**. I choose to include this category to try to account for the unique effects of *political* content in tweets. Examples of these non-political but moralized topics include fat positivity, vaping, and spanking. One limitation of this category creation is that these non-political but moralized topics might be associated with one political party more than another. Additionally, I decided to *not* include the mirror tweets to these—those that are political but non-moralized tweets—as I thought this would be a potentially a messy category given recent work on political sectarianism in America (Finkel et al., 2020) and perhaps the inability of users to see political issues as non-moralized.

Third, I included tweets that were **non-political and non-moralized**, but still controversial. I chose this category to try to account for the unique effects of heavily debated issues on Twitter that were neither political nor moralized, including the Super Bowl and controversial television shows such as RuPaul's Drag Race and The Bachelor (based on qualitative examination of a sample of tweets on this topic, informed also by websites such as Buzzfeed and Reddit). Finally, I also collected **control topic** tweets about non-confrontational topics, or ones that I would not expect to give rise to much debate or disagreement.

Again, I want to highlight that confrontation and hostility are not synonymous. The controversial topic areas (political/moralized, non-political/moralized, and non-political/non-moralized) included in the tweet collection are not necessarily defined by exhibiting hostility and toxicity. They are characterized as being confrontational and involving debate and disagreement (e.g., Should abortion be legal? Is spanking wrong? Who should win the Super Bowl?). Indeed, we see that when examining mean tweet toxicity for each individual topic (see Figure 3), there is a range of overlapping toxicity (e.g., RuPaul's Drag Race, a non-political/non-moralized topic, is more toxic than climate change, a political/moralized topic).

Political and moralized	Non-political but moralized
Abortion Black Lives Matter movement Climate change Gun control	Fat positivity Gambling Nuclear power Spanking Vaping
Non-political and non-moralized	**Control topics**
2020 Super Bowl The Bachelor Grey's Anatomy RuPaul's Drag Race Tiger King	Pets Fashion Make-up Travel Well-being

Table 10. Tweets included in this study were classified into three broad categories of controversial topics, as well as a control category. A list of the issues in each category is listed below. For a full list of phrases associated with each topic, see Appendix A.

Measures

Measuring toxicity of tweet. Tweet toxicity, the operationalization of confrontational behavior on Twitter, was measured by using the Perspective API algorithm. Perspective API was created by Jigsaw and Google's Counter Abuse Technology team through the Conversation AI research initiative.[8] Perspective API uses machine learning models based on human-labeled training datasets to classify text comments on a continuous 0-1 scale as toxic, defined as "a rude, disrespectful, or unreasonable comment that is likely to make you leave a discussion."[9] Applying Perspective API, tweets are given a continuous rating of the probability of toxicity between 0 and 1, where 1 refers to the highest probability that a tweet includes toxic language. Some researchers have found issues with the Perspective API rating process (Hosseini et al., 2017). Specifically, the algorithm sometimes classifies non-aggressive profanity, excessive punctuation,

[8]For API documentation, see https://conversationai.github.io/.
[9] https://www.perspectiveapi.com/#/home

and gender and sexual identity indicators as being more likely to be toxic (Hosseini et al., 2017; J. H. Park et al., 2018). However, Perspective API is still commonly used by highly-trafficked websites such as Wikipedia[10] and the New York Times[11] to classify toxic comments and has been used to measure hostility in recent academic research (e.g., Awal et al., 2020; Obadimu et al., 2020).

Measuring relationship closeness between Twitter users. I operationalized relationship closeness on Twitter as the shortest path network distance ("**network distance**"), which has been previously used in communications and information research to examine network connectivity on Twitter (e.g., Bakhshandeh et al., 2011; Ch'ng, 2015; Myers et al., 2014).[12] Network distance, as applied to Twitter, is a measurement of the length of the shortest path between the tweet sender and the tweet receiver; the path is the number of "friendship" hops it takes to move between the sender and the receiver.[13] Network distance is a categorical variable measured on a scale from 1 to 4 or greater ("4+"), with 1 representing a close relationship and 4+ representing a distant relationship.[14]

To illustrate network distance as a metric for relationship closeness with a more intuitive example, the network distance between one of my committee members and my sister on Twitter would be two (committee member → Susannah → Susannah's sister. On Twitter, a connection

[10] https://meta.wikimedia.org/wiki/Research:Detox
[11] https://www.nytco.com/press/the-times-is-partnering-with-jigsaw-to-expand-comment-capabilities/
[12] Initially, I used the Jaccard Index (e.g., Bindu et al., 2018; Culotta & Cutler, 2016) to measure relationship closeness. However, I ultimately decided to use network distance instead because it is a more concise measure, and it allows for measurement of network distance beyond friend-of-friends, unlike the Jaccard Index (i.e., beyond a network distance of 2).
[13] Colloquially, this effect is easier to make sense of if you know the game "Six Degrees of Kevin Bacon," in which you try to connect any Hollywood actor to Kevin Bacon in the least number of shared movies (https://oracleofbacon.org).
[14] I choose 4 or greater as the final category to maintain relative sample size balance between the categories above network distance of 1 (i.e., 50% of the category lengths were of network distance 1, and then the remaining network distances of 2, 3, and 4 + were around 16-17%.

between two people (i.e., a "friendship" hop) is defined as a reciprocal Twitter mention between a sender and a receiver. Thus, to return to the example above, this means a committee member must have directly tweeted at me, and I directly tweeted at my committee member in return (and the same for my sister and me). Even if there are longer or more indirect paths to connect two people (such as committee member → Susannah → Susannah's Mom → Susannah's sister), I am only considering the *shortest* paths as the metric for distance between two people. Operationalizing relationship closeness as this network distance metric makes the assumption that closer Twitter friends will have more conversations with common friends.

Results

Descriptive and frequency data. A total of 213,304 tweets were analyzed. The average tweet had a low probability of being toxic ($M = 0.23$, $SD = 0.20$), which makes sense given prior research that most people talking on Twitter are talking to people who hold similar beliefs, and are thus likely less toxic (e.g., Barberá et al., 2015). Indeed, supporting this claim, counts of the categorical variable of network distance indicated that tweets are more common between users who are close (and who I would expect to hold more similar beliefs); 50% (N = 108,207) of tweets had a network distance of 1, 17% (N = 35,939) had a network distance of 2, 17% (N = 35,141) had a network distance of 3, and 16% (N = 34,017) had a network distance of 4 or more.[15] The sample was skewed in terms of the number of tweets in each topic category: 87% (N = 185,553) of the tweets were control topics, 8% (N = 17,126) were political/moralized, 4% were non-political/non-moralized (N = 8,497), and 1% were non-political/moralized (N = 2,128). Also, there was a range of mean toxicity within each individual topic as well as between the four

[15] The breakdown of tweets at or beyond a network distance of 4 was 10% (N = 21,113) at network distance of 4, 4% (N= 9,238) at a network distance of 5, 1% (N = 2,770) at a network distance of 6, and <1% (N=896) for a network distance of 7 to 12.

topic categories (see Figure 3). Although the political/moralized and non-political/moralized tweets tended to be higher in mean toxicity than the non-political/non-moralized and control tweets, there was still overlap in the topics, suggesting a somewhat representative sample range in topic selection.

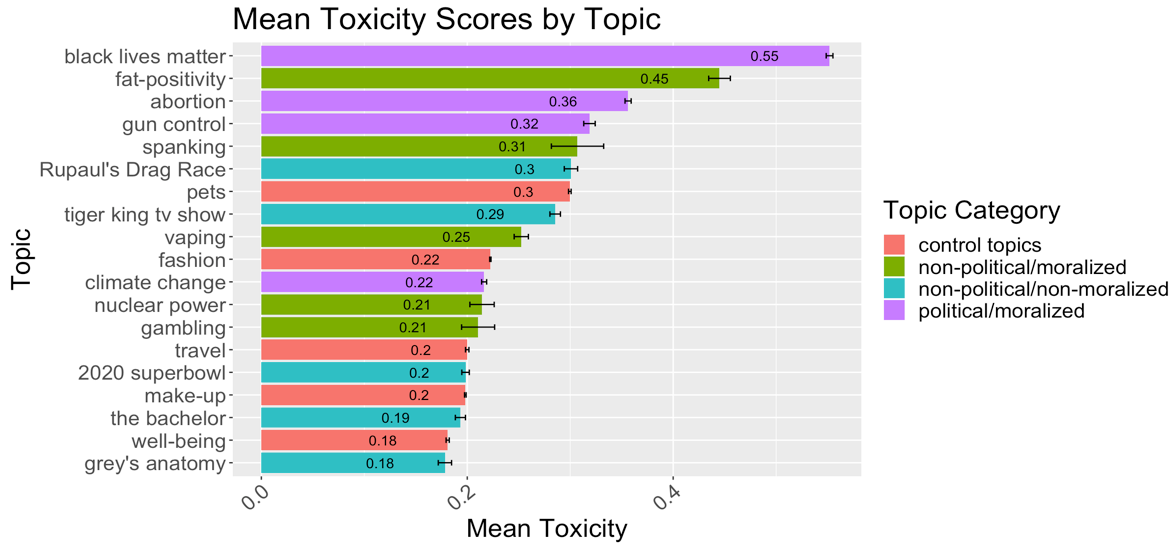

Figure 5. Twitter topics arranged by mean toxicity. Color indicates the topic category.

Multilevel models. I examined how well network distance—a metric that approximates the construct of relationship closeness—predicted the toxicity in the language of a tweet, as operationalized by Perspective API score. I used multilevel modeling to account for possible variation due to a nested data structure of tweets within users. I added a random effect for individual Twitter users, i.e., multiple tweets to the same receiving user and/or by the same sending user, in order to account for any variation that might have been due to the same users receiving multiple tweets or the same user sending multiple tweets, which occurred a total of 73 times in the dataset.

I added the following covariates in both models: account age (in years), number of followers, number of friends, number of friends for the tweet sender and receiver, and number of total tweets for both the tweet sender and receiver, as well as tweet length (in number of words).

I included these covariates because they were significantly correlated with the main outcome variable, toxicity, but I wanted to parse out the unique effects of my predictor variables of interest: network distance and tweet topic category.

Interaction effects of network distance and topic category on likelihood of tweet toxicity. There was a significant interaction between network distance and topic category, $F(3, 213275) = 13.94$, $p < .001$. Specifically, political/moralized and non-political/moralized tweets both had a higher likelihood of being toxic than tweets in either the non-political/non-moralized and control topic categories. However, only for the political/moralized tweets did the effect of network distance on likelihood of toxicity significantly *increase* in strength across increased network distances (see Figure 4). For political/moralized tweets: at length 2, B = .01, $SE = .004$, $t(213{,}300) = 3.07$, $p = .002$; at length 3, B = .03, $SE = .004$, $t(213{,}300) = 7.04$, $p < .001$; at length 4+, B = .04, $SE = .004$, $t(213{,}300) = 10.15$, $p < .001$. What this result indicates is that as social distance increases between two Twitter users, it is more likely at each increase in distance that the tweet will be toxic, for tweets that are both political and moralized. For non-political/moralized tweets, there was an intercept-level increase in the likelihood of toxicity as compared to the control and non-political/non-moralized tweets; however, the effect remained unchanged over increased distance.

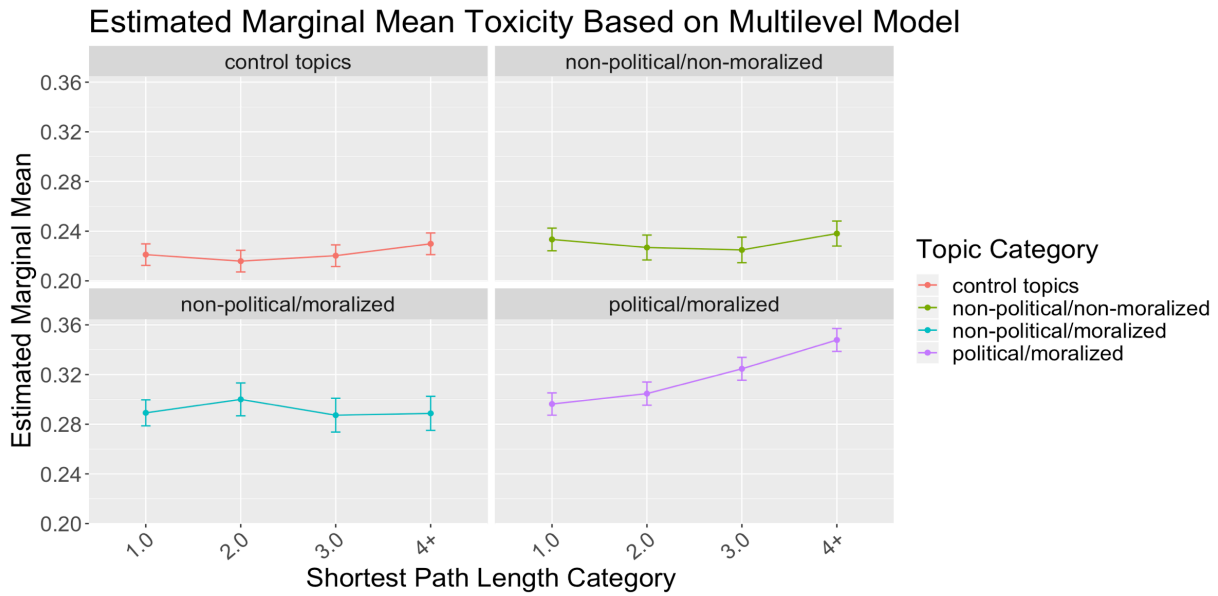

Figure 6. Marginal mean toxicity estimates for the multilevel model interaction effects between topic category and shortest path length category. Both moralized topic categories (bottom two graphs) start with a higher intercept-level estimated marginal mean toxicity.

Main effects of social distance and topic category on likelihood of tweet toxicity.

There was a significant main effect of network distance on predicted tweet toxicity, such that increases in the shortest path length between two users (a proxy for decreased relationship closeness) predicted increases in the probability that a tweet was toxic, B = 0.002, *SE* = 0.0004, $t(213{,}300) = 5.11$, $p < .001$.[16] There was also a significant main effect of topic category, $F(3, 213275) = 1194.86$, $p < .001$.

In order to further explore the effects of network distance on toxicity, I broke down the individual effects of each network distance category (from shortest path length of 2 to shortest path length of 4+, compared to the shortest path length of 1) and of each topic category (political/moralized, non-political/moralized, and non-political/non-moralized, compared to the control topics category) on the likelihood of a tweet being toxic. Please note below that negative

[16] For this multilevel model, I entered network distance as a continuous variable to measure if there was an overall linear effect of increasing social distance. Notably, when network distance is entered as a categorical variable, the differences between the categories is also significant, $F(3, 213275) = 6.88$, $p < .001$.

regression weights refer to a lower likelihood of tweet toxicity and positive regression weights refer to a higher likelihood of tweet toxicity.

There was a significant main effect of network distance at the shortest path length categories of 2 and of 4+ on the likelihood of tweet toxicity across *all* tweet topic categories (Shortest path length of 2, B = -0.005, *SE* = 0.001, $t(213{,}300) = -4.02$, $p < .001$; Shortest path length of 4+, B = 0.009, *SE* = 0.001, $t(213{,}300) = 5.596$, $p < .001$. Contrary to what I expected, this result suggests that, overall, there is a significant *decrease* in the probability of tweet toxicity as the shortest path length goes from 1 to 2, before the effect reverses and then there is a significant *increase* in the probability of tweet toxicity as network distance increases to 4+ (see Figure 7). However, when looking at the *interaction* between network distance and topic category for political/moralized tweets, there is a significant positive linear relationship between increasing network distance and increasing likelihood of tweet toxicity, as predicted.

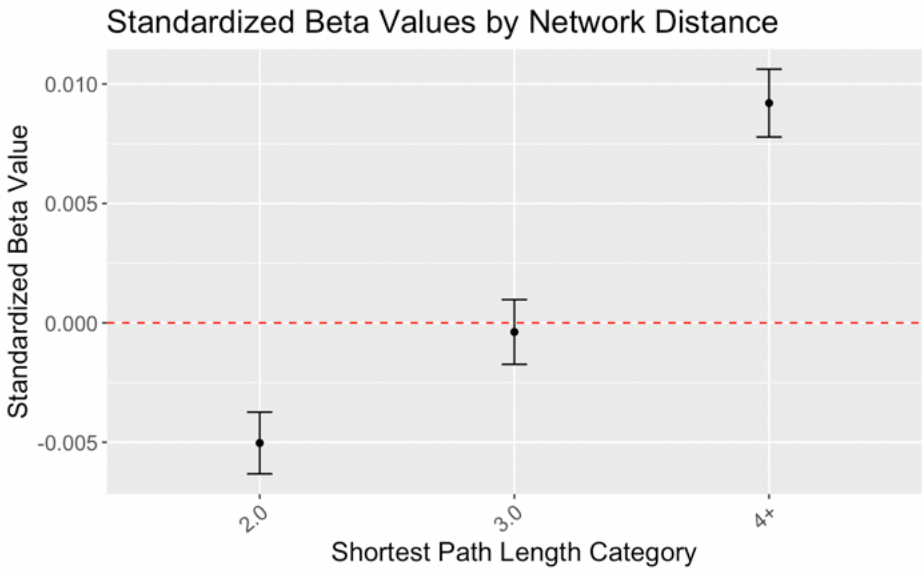

Figure 7. Plot of standardized beta values for the main effects of shortest path length category (i.e., network distance) on toxicity.

Results for the main effect of topic category showed that all three controversial topic categories (political/moralized, non-political/moralized, and non-political/non-moralized) predicted significant increases in the probability of tweet toxicity as compared to the control topic category; for political/moralized topics, B = .08, *SE* = .003, $t(213{,}300) = 29.98$, $p < .001$; for political/non-moralized topics, B = .07, *SE* = .006, $t(213{,}300) = 11.44$, $p < .001$; and for non-political/non-moralized topics, B = .01, *SE* = .003, $t(213{,}300) = 4.08$, $p < .001$. Notably, the effect is stronger for both the moralized topics than the non-moralized topics.

As a robustness check, I also performed cross-validation on the multilevel model on a random subset of 50% of the data to ensure robustness of the effects across random subsamples. Critically, I found that in a random subsample of 50% of the tweets, increased network distance still significantly predicted an increased probability of tweet toxicity, B = 0.003, *SE* = 0.0007, $t(106{,}600) = 4.45$, $p < .001$.

Discussion

People can easily quickly and conveniently share their political beliefs with others on social media. But does social media merely provide us with a new context in which we can have these conversations? Or is social media a frying pan, heating up hostility above and beyond what might occur during such a conversation offline? Using multiple sources of evidence (experimental and crowdsourcing on Twitter), I demonstrate in this chapter that not only does communication context influence these conversations, but also that the *relationship between communication partners* plays a significant role. Specifically, in the present research, I find that the further the relationship between two Twitter users, as operationalized by network distance, the more likely that a tweet sent between the two about a controversial topic will include toxic language. However, this interaction only occurred for political *and* moralized tweets (e.g.,

abortion, climate change). Moralized tweets that were non-political (e.g., spanking, vaping) were more likely to be, on average, toxic than non-moralized and non-political controversial tweets (e.g., the Super Bowl, the Bachelor) or control topic tweets (e.g., well-being, make-up). However, for moralized and non-political tweets, network distance did not have a significant interaction with this topic category on predicting increases in likelihood of tweet toxicity.

This suggests something unique about morality, as well as the combination of morality and political content, in increasing the likelihood of tweet toxicity. First, tweets in all the controversial topic categories (political/moralized, non-political/moralized, and non-political/non-moralized) were, on average, more likely to be toxic than the control tweets. However, the effects were larger for the two moralized tweet categories (political/moralized and non-political/moralized). This is in-line with findings about increased moral outrage online and increased ease in responding to morally outrageous content (Brady & Crockett, 2019; Crockett, 2017).

Research on Sacred Values Theory (Atran & Axelrod, 2008; Tetlock, 2003; Tetlock et al., 2000) can help elucidate why moralized tweets (both political and non-political) are, as a whole, higher in likelihood of toxicity. Sacred Values Theory suggests that sacred values are those moral imperatives that matter most to us. They are values we cannot make trade-offs with nor compare to some sort of material gain. Oftentimes, sacred values are topics like health, life, nature, justice, and human rights (Tetlock et al., 2000), which are reflected in the choice of moralized and political tweets topics in this data collection – abortion, the Black Lives Matters

movement, climate change, and gun control. This topics are deeply personal, and often reflect one's personal identity.[17]

Although the non-political and moralized topics do have some overlaps with these sacred values (such as vaping and health), it seems that these topics are more perhaps more amenable to trade-offs and comparisons (e.g., should there be a tax on vaping products?). I might define these sacred values as less intensely-held as compared to the political and moralized ones. Perhaps that's why while we see an intercept-level increase in toxicity for the non-political and moralized tweets, there is no interaction effect with social distance. Supporting this thought, in Weidman et al. (2019), the authors found that a self-distancing intervention to reduce the tendency of people to treat moral transgressors who were friends more leniently was only effective for the most severe crimes.

Overall, this real-world research on Twitter expands upon the earlier experimental research I conducted for my master's thesis, finding that both context (face-to-face versus social media) and relationship closeness (close versus distant) influence multiple facets of confrontation about moralized political topics. However, unlike the earlier experimental research which asked participants to consider hypothetical scenarios, the present research looks at actual Twitter conversations. Past research on decision making during moral dilemmas has shown that relying on hypothetical scenarios may inaccurately reflect true behavior, lack realism (Bauman et al., 2014), and activate different regions of the brain (Kang et al., 2011). Here, I demonstrate that the effect identified in experimental studies persists to real-life social media interactions on Twitter.

[17] In support of this, I conducted a survey on Amazon Mechanical Turk with 200 respondents, where participants rated the topics in the political/moralized category as significantly more important to their personal identity than the non-political/moralized and non-political/non-moralized categories.

A few questions arise from the present research. First, the Twitter data suggest that people might perceive, to some extent, the degrees of separation between themselves and other Twitter users. However, it's not clear what cues people may use to make determinations about network distance on Twitter and if they are more (e.g., reviewing someone's past tweets to get a sense of overlapping beliefs) or less deliberate (e.g., similarity between profile pictures, usernames). Indeed, it's possible that this process is not conscious at all. Further research aiming to understand how Twitter users make these judgments could help clarify that issue.

Second, there is an unanswered question about the possible mechanisms that may make conversations about controversial topics between close friends less hostile. One of the first potential mechanisms that I see as likely is that we are worried about the potential relationship harm that could be caused by being hostile to a friend. This is a possibility that I will be exploring in chapter 3. Notably, the harm could be more other-focused (e.g., worried about hurting your friend's feelings) or self-focused (e.g., worried about harming your own feelings). In chapter 3, I will examine these two possible mechanisms—perceived harm to the other person and perceived harm to the self—as influencing this overall effect of being less hostile to close friends.

But perhaps the mechanism is less based in fear of potential consequences. Since I will not delve into this as much in the next chapter, I will briefly note it here. From past research about face-to-face communication, we see that people are more likely to pick up on empathic cues face-to-face (Schroeder & Epley, 2015; Schroeder et al., 2017). If people are interacting with a close friend on social media, might they be likely to *imagine* the empathic paralinguistic and nonverbal cues their friend would be making since they have offline experience of these types of cues?

Indeed, one of the issues with much social media research is that it ignores that many interactions that people have online are continuations of friendship formed in-person, and that a relationship may have interactions that occur in both contexts. In cases where people are already close friends offline, research shows that social media can help maintain or strengthen these relationships (Burke & Kraut, 2014). But how is your relationship with your dentist influenced by having social media contact, as opposed to just your annual cleaning? Perhaps given your brief contact with your dentist, you are less likely to imagine how he or she may react to you commenting on their social media post.

One of the limitations of this present research is that it only considered *public*, non-anonymous social media confrontation. However, the results of this study would lead us to predict that the same effects would persist to other forms of *private* non-anonymous virtual communication, such as private email and text-messaging. Furthermore, this series of experimental and crowdsourcing studies only considered the primarily text-based social media platforms of Facebook and Twitter. Other social media platforms that are primarily picture-based such as Instagram or Snapchat should be examined as well in future research studies.

Chapter 3: Mechanisms of the Social Media Incivility to Distant Others Effect

Introduction

In the previous chapter, I presented converging evidence—from six experimental studies and one crowdsourcing study on Twitter—for a robust effect indicating that relationship closeness is inversely related to incivility on social media for moralized political topics. Having established this foundation, I now turn to exploring potential *mechanisms* that might elucidate this effect in two studies. In study 1, I will investigate how applicable six possible mechanisms are in understanding this phenomenon. Specifically, I will systematically examine participants' open-ended text responses describing their thoughts and feelings as they consider confronting another person about a moralized political topic. In study 2, I will use a similar paradigm to the previous experimental studies to test if these mechanisms can be *manipulated* to dampen or reverse the effect of decreasing relationship closeness on increased social media incivility.

Background

I propose that there are two competing motivations at play when considering the following behavior: confronting someone about their differing views on a controversial, moralized political issue. On the one hand, humans are motivated to preserve our current relationships with close others (Argyle & Henderson, 1984; Canary & Stafford, 1994) and cooperate, even with strangers, (Fehr et al., 2002; Fehr & Fischbacher, 2004). We also generally enjoy building new positive social connections (Baumeister & Leary, 1995; Rusbult & Buunk, 1993). Achieving goals consistent with these motivations can be threatened by showing

disagreement with another person, and therefore humans often dislike and avoid these situations (Milgram & Sabini, 1978; Rosen & Tesser, 1970).

On the other hand, humans are also motivated to express our beliefs to others who hold strong, but opposing, views to our own, especially about moralized political topics (Henrich & Boyd, 2001; Schein & Gray, 2016, 2018). Our desire to preserve close relationships may interfere with this motivation. If people do choose to confront a close friend about their opposing beliefs, choosing to do so in a more intimate communication context—like an in-person conversation or phone call—can foster increased understanding and reduced hostility (Kumar & Epley, 2020; Mutz, 2006). It allows us to more deeply explain our beliefs, as well as communicate nonverbally. Supporting this, my earlier experimental research indicates that choosing to confront a friend face-to-face or on the phone feels more comfortable and less effortful (Chandhok et al., in prep). But it's not clear if this choice is motivated by a desire to have a helpful, civil debate with a friend, or merely avoid harming the friendship, among other potential motivations.

A reverse situation may occur where our motivation to express our outrage interferes with the motivation to cooperate and get along when one confronts a distant other about their opposing beliefs face-to-face. In this case, when confronting a distant acquaintance face-to-face, our motivation to cooperate might supersede our motivation to express disagreement, causing internal discomfort or dissonance. However, conversing on social media seems to decrease this discomfort (Chandhok et al., in prep). A social media post allows people to more easily express their opposing beliefs, without the discomfort of being non-cooperative (perhaps because, as suggested by Joinson, 2001, people are less self-aware when disclosing online). But again, what

accounts for this asymmetry when comparing decisions about how to confront a distant versus close other?

Related research can begin to help us make sense of this asymmetry. Past work has identified multiple mechanisms for a similar effect—that people are more lenient in punishing others based on moral violations if those people are friends (Earp et al., 2020; Weidman et al., 2019). For example, Weidman et al. (2019) found that mechanisms such as self-interest, consideration of harm to the transgressor, and consideration of harm to society mediated the relationship between relational closeness and punishing a moral transgressor. However, no research has yet examined the mechanisms for the effect that relationship closeness is inversely related to incivility on social media. In the next section, I propose six potential, but non-exhaustive, mechanisms that may be relevant, based on prior research.

Potential mechanisms of the social media incivility to distant others effect

Self-relevant consequences. As a species, humans are generally motivated to seek pleasure and avoid pain. As discussed before, humans tend to feel discomfort when confronting distant acquaintances in-person (Milgram & Sabini, 1978; Sabini et al., 2000) and close friends on social media (Chandhok et al., in prep). This discomfort can be felt internally, and we may seek to relieve ourselves of this pain by either avoiding an uncomfortable situation (e.g., Rosen & Tesser, 1970; Tesser & Rosen, 1975) or, if we cannot avoid it, choosing the least effortful and most comfortable option (Chandhok et al., in prep).

Furthermore, a number of foundational ideas in social psychology, (e.g., the self-reference effect, the actor-observer bias, naïve realism) suggests that we are more able to easily recall information, attributions, and experiences when they are related to ourselves (Miller & Norman, 1975; Robinson et al., 1995; Rogers et al., 1977). Therefore, I predict that a majority of

participants (>50%) in either the close or distant conditions will report thinking about the self-relevant consequences of choosing to confront someone else on the phone or on social media during an open-ended text response (i.e., this will be the most frequent theme coded). But I don't expect there to be significant differences between the close and distant conditions, given the salience of self-relevant information.

Other-relevant consequences. Humans are social creatures, and we often consider how our actions affect other people. If we predict that confronting another person might cause them discomfort, we might consider approaching them in a less harmful way to protect their feelings. However, the "less harmful" way might be different when considering confronting a friend or distant acquaintance. We are motivated to care for people we are close to, as well as think about potential implications of our behavior on their well-being, because we see protecting close others as an extension of protecting ourselves (Aron et al., 1991). We might also seek to avoid the potential harm that could come to a friendship by confronting the other person, and instead aim to maintain the relationship (Canary & Stafford, 1994). Confronting a friend in-person or on the phone might present more opportunities to preserve the relationship. Therefore, I predict that people in the close condition will be significantly more likely to report thinking about the other-relevant consequences of choosing to confront a friend than in the distant condition.

Reputational-relevant consequences. Humans are motivated to maintain a positive standing within our in-group and cooperate with our other in-group members (Fehr & Fischbacher, 2004). We are motivated to be non-confrontational toward our in-group in order to avoid rejection and exclusion, which are both threatening to our long-term personal health (e.g., Baumeister & Leary, 1995) and can even physically hurt (Eisenberger et al., 2003). However,

these thoughts might be subsumed under self-relevant consequences, rather than made as a conscious connection to one's reputation.

On the other side, if our in-group members see us confronting an out-group member (e.g., someone who has a different view about abortion), it could actually be beneficial to signal a commitment to the in-group's beliefs by speaking out against the naysayer (Fehr et al., 2002; Fehr & Fischbacher, 2004; Henrich & Boyd, 2001). This benefit is more externally motivated by increasing support from in-group members, for example in the form of a like or comment on one's social media post, and thus more conscious as a reputational benefit. Therefore, I predict that people in the distant condition will be more likely to think about the reputation-relevant consequences of choosing to confront a distant acquaintance than those in the close friend condition.

Society-relevant consequences. Finally, the topic that I am asking people to consider when confronting another person in the present two studies—abortion—is an issue that is frequently linked to its societal implications. Research shows pro-life advocates often espouse deeply-held societal beliefs about protecting human life, while pro-choice advocates speak up about protecting a woman's right to be an autonomous agent in society (Hanschmidt et al., 2016; Norris et al., 2011). Extending upon the reputation-relevant consequences, one reason why we might be motivated to share and protect the views of our in-group is because we believe these views have consequences for society as a whole. Indeed, pro-life and pro-choice advocates often share their views in public (consider in chapter 2 that we collected over 14,334 tweets on abortion). However, while I would predict that sharing in public is more for the goal of charismatic signaling to one's in-group (Tur et al., 2021) than a desire to change the view of on out-group member that might come across this content, participants might avoid offering this

reasoning in an open-ended text response because charismatic signaling seems less authentic. Therefore, I predict that people in people in both the close and distant conditions will think about the society-relevant consequences of the decision at equal rates.

Harm and help-focus. In a pilot study, where participants were asked to describe a recent confrontation, the open-ended text comments suggested that some participants thought about the helpful consequences of engaging in a confrontation (e.g., improving the friendship by having an open discussion; improving society by sharing one's point of view) while others thought about the harmful consequences (e.g., feeling bad after confronting another person; hurting a friendship by disagreeing with their point of view). This dimension of the helpfulness/harmfulness of the consequences mapped onto the self, other, reputation, and society themes, but it seems there may be something unique about measuring this dimension. Specifically, I predicted that people would be more likely to consider the helpful consequences of confronting a friend in-person or a stranger on social media, and the harmful consequences of confronting a distant acquaintance in-person or a friend on social media, based on my earlier experimental studies suggesting that the latter situations (the more harmful ones) are both more effortful and less comfortable to imagine.

Overview

To test these possible mechanisms, I conducted two studies – one exploratory and one confirmatory – to examine the effect of relationship closeness and communication context on confrontational discussions about politics. In study 1, I asked participants to describe in words their perceptions about confronting either a close friend or a distant acquaintance on social media before choosing how to confront them. A team of research assistants, blind to the predictions and study design, coded the linguistic data to explore the four themes described above (self, other, reputation, society) as well as the help/harm dimension. In study 2, I asked people to imagine

confronting close or distant others in-person or on social media, identical to the prior manipulation from my set of six earlier experimental studies (Chandhok et al., in prep). But in study 2, I measured possible mechanisms, as informed by study 1.

Study 1

Method

Participants. I recruited 204 participants (M_{age} = 42.02, SD_{age} =13.94; 127 females; 69% White, 13% Asian, 10% African American, and 8% multiracial or other) through Amazon Mechanical Turk (MTurk) via the data collection platform TurkPrime. These crowdsourcing websites have been shown to be effective tools for obtaining high-quality data in the social sciences as compared to traditional in-person recruitment and data collection (Buhrmester et al., 2011; Litman et al., 2017). Participants were compensated $1.00 for an expected 20-minute study (although the average time was 14 minutes). To qualify for the study, participants had to be over 18 years of age and use at least one social media platform regularly. Although no participants selected "no" to our data quality question,[18] four participants did not follow the directions in writing responses to the essays (e.g.., copy/pasted the instructions, wrote "nice survey," etc.) and were excluded from subsequent analyses, bring the total sample size to 200.

Experimental Manipulation. After obtaining informed consent, participants were told they would be participating in a study examining how people express their thoughts and feelings to others. This study was hosted on Qualtrics. In the first section, participants were randomly assigned to imagine (1) someone who is very *close*, i.e., a best friend, or (2) someone who is a *distant* acquaintance. We asked participants to write the name of the person they were thinking

[18] Our data quality question was: "As researchers, the quality of our data is very important to us, so we want to make sure that your responses are valid and authentic. You will still receive payment for participating in this study, regardless of how you answer the question below. In your honest opinion, should we use your data?"

of and their relationship to them (e.g., “my best friend Jack” for the close condition or “my dentist Jill” for the distant condition).

In the second section, all participants were instructed to read and imagine the following confrontational scenario:

> *One day while reading the news, you come across an opinion article about abortion. You read the article and find that not only do you disagree with the writer's point of view about abortion, but this writer has also drawn upon flimsy and false evidence to support their claims.*

Then, I piped the text from the first section of the survey where participants wrote the name of the person they were thinking of and their relation to them into the text of the confrontational scenario (e.g., participants would read the confrontational scenario with the added line, “You discover that the writer of the article is your best friend Jack” or “You discover that the writer of the article is your dentist Jill”).

Next, all participants were asked to imagine that they decided they *will* respond to the writer of the article to share their own views about abortion. Participants were then asked to consider two possible options for responding: first, on the *phone* (e.g., “You can call Jack on the phone to explain why you disagree with their opinion article) or, second, on *social media* (e.g., You can share the opinion article on your Facebook page, along with a short post of why you disagree with it.).

Before asking participants to make a final decision, I asked them to write 3-5 sentences about what was running through their minds as they considered this decision. Specifically, I asked:

In order to better understand your thought process in making a decision about how to respond, we would like you to write down everything that you are thinking about as you contemplate this decision. In the space below, please fully explain what you're thinking in a few (3-5) sentences. Remember to fully immerse yourself in the situation and honestly articulate exactly what you are thinking and feeling.

After responding to this prompt, participants were asked to make an ultimate, binary choice decision about whether they would choose to respond to the writer of the article on the phone or on Facebook. I chose a phone call rather than in-person interaction because, in my master's thesis, I found that the pattern of results was the same whether participants thought about confronting someone in-person or on the phone (as compared to on Facebook). Given that I ran the present study during COVID-19, I decided for realism to have the response choice be phone or Facebook, since people were no longer meeting in-person.

Coding Scheme. Two research assistants, blind to the study's hypotheses and the experimental manipulation, coded the open-ended text responses based on six themes discussed in the introduction: (1) self-relevant consequences of the decision (e.g., "This would make me uncomfortable"), (2) other-relevant consequences of the decision (e.g., "This would make the person I am confronting upset"), and (3) group-relevant consequences of the decision (e.g., "This would make me look bad to other people"), (4) society-relevant consequences of the decision (e.g., "This would help society if I expressed this opinion), (5) harm-focused (e.g., "This would hurt our friendship), and (6) help-focused consequences (e.g., "We could find common

ground").[19] The research assistants coded the essays based on a binary scale (0 = "theme absent"; 1 = "theme present").[20]

The research assistants showed good agreement on each dimension (Gwet's ACs >.72), except for reputation (= .57). Given this dimension's lack of reliability, I decided not to analyze it further. I used Gwet's AC as an index of reliability instead of Cohen's kappa due to low base rates of observing each coding theme, i.e., each theme appeared in less than 50% of the essays (this is different from the number of essays that included any theme; only 8% of essays were not coded with any theme) (Gwet, 2008; Spitznagel & Helzer, 1985). For cases in which the research assistants disagreed, the code was averaged to 0.5.

Results

Are there differences in the prevalence of the themes in the close versus distant conditions?[21] The prevalence of two themes emerged as significantly different between the two relationship conditions (Table 1). First, participants who imagined confronting a close other mentioned the other-relevant consequences of this decision significantly more often (45% of essays) than participants who imagined confronting a distant acquaintance (16% of essays), B = 0.26, SE = 0.06, $t(198) = 4.46$, $p < .001$, controlling for the length of the essay (by number of words). Second, participants who imagined confronting a close other also mentioned the help-

[19] A pilot study (N = 30) indicated that the research assistants were able to reliably identify the themes. Additionally, we found evidence for help and harm themes to code for, which we added to the final coding scheme.

[20] In the pilot study, the research assistants rated each dimension on a four-point scale (0 = "no mention"; 1 = "minor theme"; 2 = "moderate theme; 3 = "major theme"). However, given that the essays were sometimes short (two or three sentences), my research assistants reported that it was difficult to assign a more granular score to each essay than either present or not present. Therefore, for the final coding scheme, participants coded the essays based on a binary scale (0 = "theme absent"; 1 = "theme present").

[21] Note that because the sample size in each condition equals 100, the mean values in each condition are also equal to the percentages of essays in that condition that included that theme (e.g., the mean for including a self-relevant consequence in the close condition is .42, and thus 42% of the close condition essays include a reference to a self-relevant consequence).

relevant consequences of this decision significantly more often (20% of essays) than participants who imagined confronting a distant acquaintance (9% of essays), B = .09, SE = 0.04, *t(*198) = 2.13, *p* = .03, again controlling for the length of the essay.

Although the other themes did not differ significantly between the relationship conditions, there was a general trend that participants in the close condition tended to more frequently describe the potential consequences of their decision to confront a friend, at least based on the dimensions we coded for (see Figure 1). This could suggest that the participants in the close condition were being more thoughtful in their essays, a suggestion also supported by the finding that participants on average wrote marginally significantly more words per comment in the close essays (M = 18.96, SD = 0.81) than the distant essays (M = 16.88, SD = 8.65), *t*(198) = 1.75, *p* = 0.08. Furthermore, the only theme where mentions of the theme were nearly the same was in the society-relevant condition.

	Close condition *N* = 100	**Distant condition *N* = 100**	**t**	**df**	***p*-value**
Self-relevant consequences	*M* = 0.42 SD = 0.45	*M* = 0.32 *SD* = 0.43	1.60	198	.11
Other-relevant consequences	*M* = 0.45 *SD* = 0.48	*M* = 0.16 *SD* = 0.35	4.46	198	< .001*
Society-relevant consequences	*M* = 0.16 *SD* = 0.32	*M* = 0.16 *SD* = 0.34	-0.22	198	.83
Harm-related consequences	*M* = 0.09 *SD* = 0.24	*M* = 0.07 *SD* = 0.17	0.67	198	.50
Help-related consequences	*M* = 0.20 *SD* = 0.36	*M* = 0.09 *SD* = 0.24	2.13	198	.03*

Table 11. Welch's t-test for presence of the themes by relationship closeness.
* p < .05

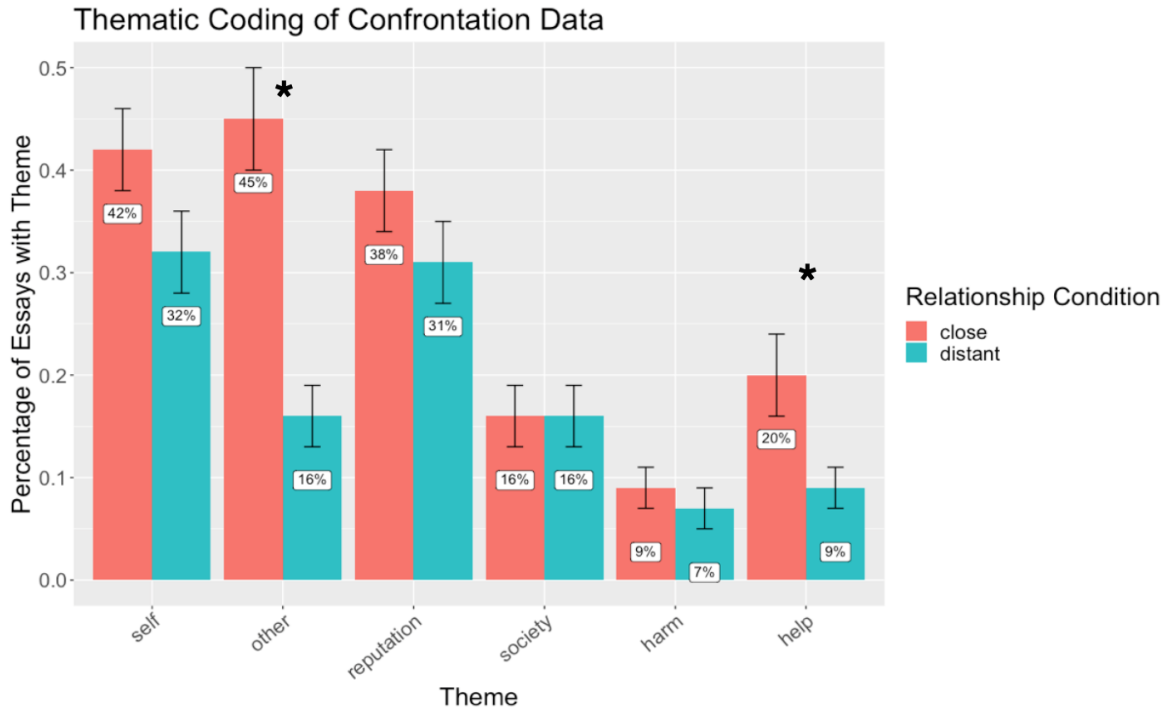

Figure 8. Percentage of essays coded with the six themes, grouped by relationship condition. * refers to a significant difference between the close and distant relationship conditions on frequency of the theme at $p < 0.05$, after controlling for the length of the essay.

Does relationship condition influence the communication context in which participants ultimately decide to respond? Using a chi-square test of independence, I examined the relationship between condition (close versus distant) and response choice (phone versus Facebook). The relationship between these variables was significant, X^2 (1, $N = 200$) = 53.65, $p < .001$. Participants in the close condition were more likely to choose to confront a friend over the phone. Specifically, 88% of participants in the *close* condition ultimately chose to confront the writer on the *phone*. On the other hand, 62% of participants in the *distant* condition ultimately chose to confront the writer on *Facebook*. There were no significant interactions between the relationship condition and participants' mentioning any of the six themes in predicting either response choice.

Discussion

Study 1 provides evidence that a potential mechanism driving people in the close condition to be likely to choose to confront a close friend over the phone is that they are more aware of the other-relevant consequences of the decision, as well as the helpful aspects of confronting a friend on the phone. Notably, the effect is largest between the close and distant conditions for the theme of other-relevant consequences. Based on this, I decided that this would be a useful mechanism to test for in Study 2. More specifically, I predicted that if participants in the distant condition were encouraged to think about the other-relevant consequences of their choice to either confront a distant acquaintance on the phone or on Facebook, these participants would be more likely to choose the phone call. I hypothesized that this other focus would perhaps encourage participants to be more empathetic to the distant acquaintance they are confronting, and therefore want to make the confrontation more comfortable for them.

I also wanted to examine if there was a condition, based on the thematic coding, under which participants in the close condition might be more likely to confront a close friend on Facebook. The best candidate for that condition, based on the data, would be the society-relevant consequences. I predicted that if participants in the close condition were encouraged to think about the society-relevant consequences of their choice, they might be more likely to decide to post on Facebook. For example, one quote that stood out to me in the open-ended text responses in study 1 was a participant, in the distant condition, who said they would want to post on Facebook "as a public service." This led me to consider that if participants in the close condition were encouraged to think about how sharing their views might influence societal views about abortion in their favor, they might be more encouraged to post this opinion on Facebook. In Study 2, I was able to examine these predictions.

Study 2

Method

Participants. I recruited 200 participants (M_{age} = 41.80, SD_{age} =13.44; 135 females; 82% White, 9% African American, 5% Asian, and 4% multiracial or other) through Amazon Mechanical Turk (MTurk) via the data collection platform TurkPrime. Participants were compensated $1.00 for an expected 20-minute study (although the average time was 9 minutes). To qualify for the study, participants had to be over 18 years of age and use at least one social media platform regularly. No participants initially were excluded from the study for not following directions at all or selecting "no" to our data quality question.

Experimental Manipulation. Using nearly the same paradigm as study 1, participants were first randomly assigned to imagine either a close friend or a distant acquaintance. We asked participants to write the name of the person they were thinking of and their relationship to them. Then, participants read the same confrontational scenario as in study 1 with the piped text of either their close friend or distant acquaintance.

However, before making a decision about how to respond, participants were asked to think about and write two to three sentences about how their choice (to call or post on Facebook) would affect either (1) the writer of the article (i.e., their mood, their self-esteem) or (2) society in general (i.e., the spread of misinformation, public perceptions of the topic). This manipulation condition (mapping onto either an other-focus or society-focus, respectively) was based on the results for Study 1, which indicated people in the close condition are more likely to think about other-relevant consequences of how they choose to respond, while people in the distant condition are about equally likely to think about society-relevant consequences as in the close condition (which was the most balanced theme between the two relationship conditions).

After the experimental manipulation, participants were asked to make a binary choice about how they would ultimately choose to respond to the writer of the op-ed—either over the phone or on Facebook.

Results

Does relationship closeness influence choosing to confront another person on Facebook? A logistic regression indicated a significant effect of relationship closeness on the choice to confront another person on Facebook, such that the odds of choosing Facebook are 29 times greater if considering confronting a distant acquaintance (B = 3.37, SE = 0.66, z = 5.08, p <.001, OR = 29.04). This finding offers additional evidence to support the finding from study 1 that people in the distant condition are more likely to prefer confronting a distant acquaintance on Facebook. Neither the manipulation, i.e., writing an other-focused essay (B = 0.55, SE = 0.76, z = 0.73, p = .47, OR = 1.74), nor of the interaction between being in the distant condition and being asked to write an other-focused essay (B = -0.74, SE = 0.87, z = -0.85, p = .39, OR = 0.48) significantly influenced response choice.

Does relationship closeness influence choosing to confront another person on the phone? A second logistic regression showed a significant effect of relationship closeness on the choice to confront another person on the phone, such that the odds of choosing a phone call are almost 14 times greater if considering confronting a close friend versus a distant acquaintance, (B = 2.63, SE = 0.56, z = 13.85, p < .001, OR = 13.85). Again, neither the manipulation, i.e., writing a society-focused essay (B = -0.19 SE = 0.42, z = -0.45, p = .65, OR = 0.83), nor the interaction between being in the close condition and being asked to write a society-focused essay significantly changed the results, (B = 0.74, SE = 0.87, z = 0.85, p = .39, OR = 2.10).

Predictor	B	SE B	Odds Ratio	z-value	*p*-value
Response choice: Facebook					
Relationship condition: distant	3.37	0.66	29.04	5.08	<.001*
Manipulation: other-focused essay	0.55	0.76	1.74	0.73	.47
Interaction: distant x other-focused essay	-0.74	0.87	0.48	-0.85	.39

Table 12. Summary of logistic regression analysis for relationship condition and manipulation predicting the response choices of confronting the writer on Facebook. Covariates are essay length, age, gender, education, income, and ethnicity, none of which were significant and are therefore excluded from the table.

Predictor	B	SE B	Odds Ratio	z-value	*p*-value
Response choice: Phone Call					
Relationship condition: close	2.63	0.56	13.85	4.72	<.001*
Manipulation: society-focused essay	-0.19	0.42	0.83	-0.45	.65
Interaction: close x society-focused essay	0.74	0.87	2.10	0.85	.39

Table 13. Summary of logistic regression analysis for relationship condition and manipulation predicting the response choices of confronting the writer with a phone call. Covariates are essay length, age, gender, education, income, and ethnicity, none of which were significant and are therefore excluded from the table.

Discussion

This set of results suggests that the relationship between the participant and the writer of the op-ed influences how the participant chooses how to respond to the writer. However, prompting participants to consider different consequences (either other or society-relevant) of their choice of how to respond did not influence results. Either the relationship condition was so salient as to override the effects of the other versus society manipulation, or the manipulation was not sufficiently powerful. To examine the latter possibility, I went back over the data and

removed 54 data points where the respondents did not explicitly follow instructions (i.e., write about the other-relevant or society-relevant consequences of their decision). I was strict in excluding these 54 data points, looking for clear (rather than implied) linguistic indications that the participants understood the directions. However, removing these data points did not change the pattern of results. This could suggest that the manipulation (i.e., writing the other-relevant or society-relevant essay) was still not strong enough, or that relationship condition is highly relevant in participants' minds when thinking about confrontation.

General Discussion

Across two studies, relationship closeness influenced people's choices about how to confront others. Specifically, people were significantly more likely to choose to confront a distant acquaintance on social media and a close friend over the phone. Study 1 provided initial evidence of potential mechanisms of this effect. Specifically, participants thinking about confronting a close friend were more likely to think about the other-relevant and help-relevant consequences of the confrontation than participants thinking about a distant acquaintance. Given the value of close friendships (Aron et al., 1991) and the effort we put into cultivating them (Canary & Stafford, 1994), thinking through these consquences might benefit continued friendship maitenance.

In Study 2, I hypothesized that I could increase the likelihood that participants would choose to confront a distant acquaintance on the phone if they thought and wrote about the other-relevant consequences of their choice. However, I did not find significant evidence supporting that prediction. Nonetheless, given the growing evidence that online confrontations skew toward increased hostility (e.g., Cheng et al., 2017), it is important to continue examining when we can nudge people to choose a communication medium in which it might be more effortful (e.g.,

Chandhok et al., in prep), and also less likely (e.g., Mutz, 2006), to be hostile. Additionally, perhaps thinking about the other-relevant consequences of confronting a distant acquaintance could promote more civility in online spaces.

Relatedly, one of the limitations of Chapters 2 and 3 is that we did not collect linguistic data on people's *actual responses* if they had been asked to imagine confronting either a close or distant other on the phone or on Facebook. In future studies, adding this step would allow us to draw inferences about the effect of different manipulations (such as asking people to think about other-relevant consequences) on the actual hostility in their responses. Indeed, Facebook has recently experimented with an initiative asking users to rethink posting news articles if the articles are older than 90 days in order to combat misinformation spread.[22] Perhaps Facebook or Twitter could prompt users to reflect on the other-relevant consequences of sharing a post, if the post contains a certain level of hostility (which could be accomplished using Perspective API).

The findings from Chapter 3 extend my prior research on this topic in two critical progressions. First, in these two studies, I gave participants the *choice* of what communication context to use to confront either a close friend or distant acquaintance. In line with the experimental studies in my master's thesis (where I asked participants to think hypothetically about the situation where the choice was already made), participants were more likely to choose to confront a friend over the phone and a distant acquaintance on Facebook.

Second, with the linguistic data, I was able to begin unpacking the potential mechanisms driving the online incivility to distant others effect, using participants' *own* words. Incorporating and analyzing open-ended text responses is beneficial in elucidating novel psychological phenomena (such as the effect examined here), especially when supplemented by closed-ended

[22] https://www.newsweek.com/facebook-notification-old-misleading-news-warning-prompt-explained-john-hegeman-1513607

survey questions (Geer, 1988; Kjell et al., 2019). Not only do open-ended text responses allow researchers the opportunity to more fully understand our participants' subjective thoughts and feelings (Krosnick, 1999), but they are also associated with less social desirability in responding as compared to closed-ended questions (Kjell et al., 2019). Social desirability is a problem I had come across in earlier experiments studying confrontation. Finally, despite some skepticism, research suggests that participants are able to accurately articulate their attitudes in free responses (Geer, 1988).

In Chapter 2, I suggested that if people imagine confronting close friends, either in-person or on social media, they might be more likely to think through the consequences of such an action. Indeed, the essays in the close condition from Study 1 had significantly more words per essay. Perhaps interacting with a close friend in a confrontational scenario produces increased introspection, which makes sense given that people generally value and put effort into maintaining close friendships (Argyle et al., 1985; Argyle & Henderson, 1984; Canary & Stafford, 1994). However, despite this significant increase in number of words in the close condition in Study 1, the length of the essay did not significantly predict response choice. Furthermore, in Study 2, there was no significant difference in the essays' number of words between the close and distant conditions, nor between the manipulation conditions (i.e., thinking about other- or society-relevant consequences).

Conclusion

I opened this dissertation by drawing attention to the concerns that people have raised about the psychological effects of billions of people spending multiple hours per day on virtual social media platforms. The goal of my dissertation research was to compare a slice of social behaviors—self-disclosure, discussion of controversial topics, and confrontation—between face-to-face and social media contexts, aiming to elucidate the different psychological variables that influence types of interactions and their subsequent psychological outcomes. By doing this, my hope is to add nuance to the literature on the psychological effects of spending time on social media. Based on my review of the relevant literature and own dissertation research, my overall takeaway from this work is that virtual social media interactions are nuanced and complex, just as are social interactions in the physical world. Thus, I disagree with the broad criticisms I cited in the Introduction, and caution against drawing binary good-or-bad conclusions about the effects of social media use on psychological well-being. My research suggests that not all social media interactions have negative psychological outcomes.

Yet, I acknowledge the risk of ignoring the potential destructive effects of social media on well-being when they are present. In their 2010 book, *Merchants of Doubt*, historians of science Naomi Oreskes and Erik M. Conway describe how industry and corporate scientists sowed the seeds of denial and uncertainty when it came to research on the harmful effects of global warming and smoking. Scientific research is predicated on probability of uncertainty. As the authors describe, "Doubt is crucial to science…but it makes science vulnerable to

misrepresentation, because it is easy to take uncertainties out of context and create the impression that *everything* is unresolved" (Oreskes & Conway, 2010, p. 34).

The goal of this dissertation is to not undervalue people's concerns, past research, or the uncertainty inherent in science that comes with growing research in a relatively nascent field—the psychology of social media use. What I do suggest is that we cannot make sense of social media use as simply good or bad, because social media use does not have a binary outcome. Engaging in interpersonal interactions on social media is a dynamic process, and we are still fleshing out our understanding of the complex social processes that define interactions online. The goal of my dissertation is to add a block in building this knowledge.

Specifically, in Chapter 1, I compared the psychological outcomes of self-disclosures made face-to-face versus on social media. Given the positive effects of self-disclosure on social connection in face-to-face contexts (e.g., Collins & Miller, 1994, Laurenceau et al., 1998), I asked whether social media self-disclosures confer the same benefits. While I found additional evidence that face-to-face self-disclosures indeed increase feelings of social connection and affective well-being over time, the same did not hold true across the board for social media self-disclosures. Only when participants specifically were motivated to share in order *to feel good* did disclosures made on social media lead to a significant increase in affective well-being and social connection over time.

As a whole, Chapter 1 highlights that communication context is a key factor in understanding different types of social behavior. Findings from the research suggest that communicating on social media may *not* be a wholly sufficient alternative for face-to-face interactions, in terms of positive affective and social connectedness outcomes. However, this

research also suggests that self-disclosure on social media *can* be beneficial, as long as one is strategic in their motivation for disclosing (i.e., to feel good).

In Chapter 2, I turned to examining how dyads discuss controversial issues on Twitter. This work was motivated by prior literature suggesting that people are both motivated to preserve close relationships and cooperate with others, even strangers (Argyle & Henderson, 1984; Baumeister & Leary, 1995; Canary & Stafford, 1994; Fehr et al., 2002; Fehr & Fischbacher, 2004), while at the same time being motivated to confront others who hold strong, but opposing, views to our own about moralized political topics (Henrich & Boyd, 2001; Schein & Gray, 2016, 2018). My master's thesis helped reconcile how people may respond when these motivations collide. In six experiments for my master's thesis, I found that people prefer to discuss these political topics face-to-face or on the phone to a close friend, but on social media to a distant acquaintance.

In Chapter 2, I further explored this finding by collecting real-world Twitter data to see how relationship closeness influenced hostility during conversations about moralized political topics. From the Twitter data, I uncovered that the further the relationship distance between two Twitter users, as operationalized by network distance, the more likely that a tweet sent from one to the other about a *political and moralized* controversial topic (like abortion or climate change) will include toxic language.

Notably, only for the political/moralized tweets did the effect of network distance on likelihood of tweets toxicity significantly *increase* in strength across increased network distances. For non-political/moralized tweets, there was an intercept-level increase in the likelihood of toxicity as compared to the control and non-political/non-moralized tweets. However, the effect remained unchanged (i.e., did not strengthen as the effect did for

political/moralized tweets) over increased network distance. For non-political/non-moralized tweets and control topic tweets, there was no significant effect of network distance on likelihood of toxicity. I hypothesize that *political* and *moralized* controversial topics may be the most difficult for people to make trade-offs and comparisons with, as suggested by Sacred Values Theory (Tetlock et al., 2000), which is why the effect is unique to this category of tweets.

In Chapter 3, I further extended my findings from Chapter 2 in order to further examine *why* relationship closeness influences confrontational behavior, and if this effect can be manipulated. In study 1, when I asked participants to describe their thought process when confronting a close friend, they were more likely to write about how this interaction might affect their friend and their mental and emotional state. On the other hand, participants when asked to describe confronting a distant acquaintance were *slightly*—although not significantly—more likely to write about this interaction might affect society.

In study 2 of Chapter 3, I attempted to manipulate these two features in order to see if I could influence how people ultimately chose to respond to their communication partner. However, the manipulation did not significantly influence in which communication context people chose to respond. Notably, relationship closeness did significantly predict, on its own, in which communication context people chose to respond. Indeed, relationship closeness has a robust effect on the choice of context in which people desire to confront another person: a majority of people in the close condition ultimately chose to confront the close friend on the phone, and a majority of people in the distant condition ultimately chose to confront the distant acquaintance on social media.

This set of results suggests that social media interactions are indeed unique as compared to face-to-face interactions, at least when it comes to social behaviors such as self-disclosure,

discussion of controversial topics, and confrontation. The different variables that impact these social behaviors face-to-face and on social media also lead to divergent psychological outcomes.

Remaining Questions

Relative to the long tenure of the field of social psychology, the subfield of the psychology of social media use is young. Thus, after completing this dissertation research and reviewing related literature in the field, I found that many questions remained. Below, I summarize three common themes that unite some of the questions that persist in the field and in my mind, while recognizing that the below list is not exhaustive.

First, online spaces are rapidly evolving. The landscape of virtual social media interactions may not be the same when a researcher begins a research study as when she completes it. Even defining what "counts" as a social media platform can be a complex undertaking (e.g., Boyd & Ellison, 2007; Ellison & Boyd, 2013). Therefore, I think a big question that will likely persist in the field is: **how do psychological effects present on one social media platform generalize to other platforms?**

In my dissertation research, I examined people's behavior on Facebook and Twitter. However, usage of platforms like Instagram, TikTok, and Snapchat continues to rise, especially among younger populations.[23] These new platforms present a potential significant moderator for past research findings in the field. My recommendation is that perhaps the field would benefit from focusing *less* on specific social media platforms under the umbrella of a single company, and *more* on what *unites* similar platforms. For example, does the platform actively elicit users to self-disclose in a public manner (e.g., Facebook and Twitter) or in a more direct, private manner (e.g., Snapchat)? Is the primary form of social interactions text, photo, or video? Are the

[23] https://www.statista.com/topics/1164/social-networks/

interactions live or asynchronous? How much overlap is there between a user's network of social media friends and followers versus their offline social network? By focusing on categories of similar social media platforms rather than single platforms, I hypothesize that researchers might find it easier to generalize results (or vocalize why results from other studies are less applicable).

Second, many social interactions take place *across* multiple communication contexts – spanning both face-to-face and virtual worlds. I might have a conversation with a friend in-person at a coffee shop that I continue over Facebook that evening and then pick-up again in-person at the office the following day. How does this conversation look or feel different as opposed to if it took place wholly in-person or on social media? One of the affordances of social media is the capability for humans to continue having conversations with one another, despite physical distance. However, research (such as my own dissertation) sometimes takes a narrow focus on social behavior *within* one communication context rather than uncovering how social behavior might vary as one *switches* between communication contexts.

Thus, a second big question that I believe will persist in the future is: **what are the effects of *switching* between communication contexts (i.e., face-to-face versus on social media) on social interactions?** This may be a difficult research question to tackle – potential hurdles that come to mind include identifying the right "dosage" of face-to-face and social media interactions to include in an experiment, as well as how to test whether or not people switch between communication contexts based on necessity or preference. However, I ultimately believe that such research might be more ecologically valid than research that takes place only within one context.

Finally, social psychology has largely focused on the relationships people have with other humans. All three chapters in my dissertation suppose that one's communication is directed at

another human being. But as virtual technologies and social media platforms continue to develop, humans will more frequently have conversations with autonomous agents, such artificial intelligence, robots, and intelligent virtual assistants. A third big question that I believe remains in the field is: **what are the psychosocial consequences of communication between a human and an autonomous intelligent agent?** Some research exists on this topic, finding that humans will indeed self-disclose to a robot (Ling & Björling, 2020), which is perhaps unsurprising considering how rewarding humans find self-disclosure in general (Tamir & Mitchell, 2012). Indeed, humans may even engage in hostile conversations with artificial intelligence, as evidenced by Microsoft's artificially intelligent chatbox on Twitter named "Tay," who began issuing controversial tweets after engaging in machine learning from other conversations on Twitter.[24] In general, I think a key question for the future will be if humans can have positive, fulfilling relationships with autonomous agents, similar to my research question in Chapter 1 of if social media interactions were a sufficient alternative to face-to-face interactions. My hypothesis, as with Chapter 1, is that these types of interactions can be meaningful and lead to positive psychological outcomes, given intentionality and a goal for the interaction.

Concluding Comment

Humankind's ubiquitous adoption of social media means there is a new virtual world in which social behavior can occur. Does social media, as a communication context, have unique social and psychological effects? As a whole, the findings from this dissertation suggest the answer is yes—people behave differently face-to-face versus on social media when it comes to social behaviors such as self-disclosure, discussion of controversial topics, and confrontation.

[24] https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter

But notably, these effects of communication context are not so colossal as to negate the influence of other moderators, such as one's motivation for disclosure on social media or one's relationship to the person they are interacting with on social media. In an era in which concerns are mounting about how social media use affects our lives, this dissertation suggests that such nuance is fundamental.

Appendix

Full List of Phrases Included in Twitter Data Collection

Political and moralized	Non-political but moralized
Abortion	Addicted to gambling
Anti-abortion	Anti-fat bias
Black lives matter	Anti-nuclear lobby
Blm	Anti-vaping
Breonna Taylor	Body positivity
Climate action	Child physical discipline
Climate change	Compulsive gambling
Climate crisis	Corporal punishment
Climate deniers	E-cigarettes
Firearm control	Fat acceptance
Firearm deaths	Fat positivity
Firearm safety	Fat shaming
George Floyd	Gamblers anonymous
Global warming	Gambling addiction
Green new deal	Gambling disorder
Gun control	Nuclear energy
Gun deaths	Nuclear plants
Gun injuries	Nuclear power
Gun killings	Nuclear power regulation
Gun registration	Pathological gambling
Gun restrictions	Problem gambling
Gun safety	Spanking a child
Gun violence	Spare the rod and spoil the child
March for our lives	Vape juice
Mass shooting	Vape life
Nra	Vape pens
Pro-choice	Vaping
Pro-life	Vaping cdc
Reproductive rights	Vaping death
Right to choose	Vaping epidemic
Roe vs wade	Vaping lung
Save the planet	Whooping a child
School shooting	
Severe weather	
Sustainability	
Unborn babies	
Weather trends	

White privilege White supremacists	
Non-political and non-moralized	**Control topics**
49ers Alex Karev America's next drag superstar Bachelor nation Bachelorette Big cat rescue Carole Baskin Carole killed her husband Chiefs Drag race Football world champions Grey's anatomy Hard rock stadium Izzie Stevens Jimmy Garoppolo Joe exotic Lombardi trophy Meredith grey Patrick Mahomes Rpdr Rupaul Rupaul's drag race Superbowl Superbowl commercials The Bachelor The Bachelor abc Tiger King Tiger King netflix Wynnewood Exotic Animal Park	Adventures with my dog Adventures with my cat Animals Beauty Cat lover Cat of the day Cats of twitter Couture Dog lover Dog of the day Dogs of twitter Fashion Fashionista Good vibes I love my cat I love my dog Look of the day Make-up Make-up artist Make-up of the day Make-up tips Make-up tutorial Monday motivation Outfit Outfit inspiration Pets Pets' corner Self-compassion Self-love Self-care Skin care Style Travel blog Travel Tuesday Vacation Wednesday wisdom Well-being What I wore

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Acquisti, A., & Gross, R. (2006). Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In G. Danezis & P. Golle (Eds.), *Privacy Enhancing Technologies* (Vol. 4258, pp. 36–58). Springer Berlin Heidelberg. https://doi.org/10.1007/11957454_3

Agger, B. (2015). *Oversharing: Presentations of self in the Internet age*. Routledge.

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.

Andalibi, N., Haimson, O. L., Choudhury, M. D., & Forte, A. (2018). Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media. *ACM Transactions on Computer-Human Interaction*, *25*(5), 1–35. https://doi.org/10.1145/3234942

Argyle, M., & Henderson, M. (1984). The Rules of Friendship. *Journal of Social and Personal Relationships*, *1*(2), 211–237. https://doi.org/10.1177/0265407584012005

Argyle, M., Henderson, M., & Furnham, A. (1985). The rules of social relationships. *British Journal of Social Psychology*, *24*(2), 125–139. https://doi.org/10.1111/j.2044-8309.1985.tb00671.x

Aron, A., Aron, E. N., Tudor, M., & Nelson, G. (1991). Close relationships as including other in the self. *Journal of Personality and Social Psychology*, *60*(2), 241–253. https://doi.org/10.1037/0022-3514.60.2.241

Atran, S., & Axelrod, R. (2008). Reframing Sacred Values. *Negotiation Journal*, *24*(3), 221–246. https://doi.org/10.1111/j.1571-9979.2008.00182.x

Awal, M. R., Cao, R., Mitrovic, S., & Lee, R. K.-W. (2020). On Analyzing Antisocial Behaviors Amid COVID-19 Pandemic. *ArXiv:2007.10712 [Cs]*. http://arxiv.org/abs/2007.10712

Bakhshandeh, R., Samadi, M., Azimifar, Z., & Schaeffer, J. (2011). Degrees of Separation in Social Networks. *Proceedings, The Fourth International Symposium on Combinatorial Search*, 6.

Balani, S., & De Choudhury, M. (2015). Detecting and Characterizing Mental Health Related Self-Disclosure in Social Media. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, 1373–1378. https://doi.org/10.1145/2702613.2732733

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, *23*(1), 76–91.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, *26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Barnidge, M. (2017). Exposure to Political Disagreement in Social Media Versus Face-to-Face and Anonymous Online Settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554. https://doi.org/10.1111/spc3.12131

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*(3), 497–529. https://doi.org/10.1037/0033-2909.117.3.497

Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., & Falk, E. B. (2016). Sharing the small moments: Ephemeral social interaction on Snapchat. *Information, Communication & Society*, *19*(7), 956–977. https://doi.org/10.1080/1369118X.2015.1084349

Bazarova, N. N., & Choi, Y. H. (2014). Self-Disclosure in Social Media: Extending the Functional Approach to Disclosure Motivations and Characteristics on Social Network Sites. *Journal of Communication*, *64*(4), 635–657. https://doi.org/10.1111/jcom.12106

Bazarova, N. N., Choi, Y. H., Schwanda Sosik, V., Cosley, D., & Whitlock, J. (2015). *Social Sharing of Emotions on Facebook: Channel Differences, Satisfaction, and Replies*. 154–164. https://doi.org/10.1145/2675133.2675297

Bindu, P. V., Mishra, R., & Thilagam, P. S. (2018). Discovering spammer communities in twitter. *Journal of Intelligent Information Systems*, *51*(3), 503–527. https://doi.org/10.1007/s10844-017-0494-z

Bogen, K. W., Bleiweiss, K. K., Leach, N. R., & Orchowski, L. M. (2019). #MeToo: Disclosure and Response to Sexual Victimization on Twitter. *Journal of Interpersonal Violence*, 0886260519851211. https://doi.org/10.1177/0886260519851211

Bohns, V. K. (2016). (Mis)Understanding Our Influence Over Others: A Review of the Underestimation-of-Compliance Effect. *Current Directions in Psychological Science*, *25*(2), 119–123. https://doi.org/10.1177/0963721415628011

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brady, W. J., & Crockett, M. J. (2019). How Effective Is Online Outrage? *Trends in Cognitive Sciences*, *23*(2), 79–80. https://doi.org/10.1016/j.tics.2018.11.004

Brown, J. D. (1986). Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments. *Social Cognition*, *4*(4), 353–376. https://doi.org/10.1521/soco.1986.4.4.353

Bruk, A., Scholl, S. G., & Bless, H. (2018). Beautiful mess effect: Self–other differences in evaluation of showing vulnerability. *Journal of Personality and Social Psychology*, *115*(2), 192–205. https://doi.org/10.1037/pspa0000120

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1177/1745691610393980

Burke, M., & Kraut, R. (2013). *Using facebook after losing a job: Differential benefits of strong and weak ties*. 11.

Burke, M., & Kraut, R. E. (2016). The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication*, *21*(4), 265–281. https://doi.org/10.1111/jcc4.12162

Burke, M., & Kraut, R. E. (2014). *Growing closer on facebook: Changes in tie strength through social network site use*. 4187–4196. https://doi.org/10.1145/2556288.2557094

Burke, M., Kraut, R., & Marlow, C. (2011). Social capital on facebook: Differentiating uses and users. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11*, 571. https://doi.org/10.1145/1978942.1979023

Butzel, J. S., & Ryan, R. M. (1997). The Dynamics of Volitional Reliance. In G. R. Pierce, B. Lakey, I. G. Sarason, & B. R. Sarason (Eds.), *Sourcebook of Social Support and Personality* (pp. 49–67). Springer US. https://doi.org/10.1007/978-1-4899-1843-7_3

Cacioppo, J. T., & Patrick, W. (2008). *Loneliness: Human Nature and the Need for Social Connection*. W. W. Norton & Company.

Calmes, C. A., & Roberts, J. E. (2008). Rumination in Interpersonal Relationships: Does Co-rumination Explain Gender Differences in Emotional Distress and Relationship Satisfaction Among College Students? *Cognitive Therapy and Research*, *32*(4), 577–590. https://doi.org/10.1007/s10608-008-9200-3

Cameron, J. J., Holmes, J. G., & Vorauer, J. D. (2009). When self-disclosure goes awry: Negative consequences of revealing personal failures for lower self-esteem individuals. *Journal of Experimental Social Psychology*, *45*(1), 217–222. https://doi.org/10.1016/j.jesp.2008.09.009

Canary, D. J., & Stafford, L. (1994). Maintaining relationships through strategic and routine interaction. In *Communication and relational maintenance* (pp. 3–22). Academic Press.

Canevello, A., & Crocker, J. (2010). Creating Good Relationships: Responsiveness, Relationship Quality, and Interpersonal Goals. *Journal of Personality and Social Psychology*, *99*(1), 78–106. https://doi.org/10.1037/a0018186

Carr, C. T., Schrock, D. B., & Dauterman, P. (2012). Speech Acts Within Facebook Status Messages. *Journal of Language and Social Psychology*, *31*(2), 176–196. https://doi.org/10.1177/0261927X12438535

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. The Guilford Press.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone Can Become a Troll: Causes of Trolling Behavior in Online Discussions. *CSCW : Proceedings of the Conference on Computer-Supported Cooperative Work. Conference on Computer-Supported Cooperative Work*, *2017*, 1217–1230. https://doi.org/10.1145/2998181.2998213

Ch'ng, E. (2015). Local Interactions and the Emergence of a Twitter Small-World Network. *ArXiv:1508.03594 [Physics]*. http://arxiv.org/abs/1508.03594

Choi, M., & Toma, C. L. (2014). Social sharing through interpersonal media: Patterns and effects on emotional well-being. *Computers in Human Behavior*, *36*, 530–541. https://doi.org/10.1016/j.chb.2014.04.026

Choi, Y. H., & Bazarova, N. N. (2015). Self-Disclosure Characteristics and Motivations in Social Media: Extending the Functional Model to Multiple Social Network Sites. *Human Communication Research*, *41*(4), 480–500. https://doi.org/10.1111/hcre.12053

Clark-Gordon, C. V., Bowman, N. D., Goodboy, A. K., & Wright, A. (2019). Anonymity and Online Self-Disclosure: A Meta-Analysis. *Communication Reports*, *32*(2), 98–111. https://doi.org/10.1080/08934215.2019.1607516

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and Uncivil? Patterns and Determinants of Incivility in Newspaper Website Comments. *Journal of Communication*, *64*(4), 658–679. https://doi.org/10.1111/jcom.12104

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*.

Cozby, P. C. (1973). Self-disclosure: A literature review. *Psychological Bulletin*, *79*(2), 73–91. https://doi.org/10.1037/h0033950

Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, *1*(11), 769–771. https://doi.org/10.1038/s41562-017-0213-3

Csikszentmihalyi, M., & Larson, R. (2014). Validity and Reliability of the Experience-Sampling Method. In M. Csikszentmihalyi (Ed.), *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi* (pp. 35–54). Springer Netherlands. https://doi.org/10.1007/978-94-017-9088-8_3

Culotta, A., & Cutler, J. (2016). Mining Brand Perceptions from Twitter Social Networks. *Marketing Science*, *35*(3), 343–362. https://doi.org/10.1287/mksc.2015.0968

Czopp, A. M., & Monteith, M. J. (2003). Confronting Prejudice (Literally): Reactions to Confrontations of Racial and Gender Bias. *Personality and Social Psychology Bulletin*, *29*(4), 532–544. https://doi.org/10.1177/0146167202250923

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology*, *90*(5), 784–803. https://doi.org/10.1037/0022-3514.90.5.784

De Choudhury, M., & De, S. (2014). Mental Health Discourse on reddit: Self-disclosure, Social Support, and Anonymity. *Eighth International AAAI Conference on Weblogs and Social Media*, 10.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday Lies in Close and Casual Relationships. *Journal of Personality and Social Psychology*, *74*(1), 63.

Deters, F. große, & Mehl, M. R. (2013). Does Posting Facebook Status Updates Increase or Decrease Loneliness? An Online Social Networking Experiment. *Social Psychological and Personality Science*, *4*(5), 579–586. https://doi.org/10.1177/1948550612469233

Diener, E., & Seligman, M. E. P. (2002). Very Happy People. *Psychological Science*, *13*(1), 81–84. https://doi.org/10.1111/1467-9280.00415

Dunbar, R. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature*, *8*(3), 231–246. https://doi.org/10.1007/BF02912493

Earp, B. D., McLoughlin, K., Monrad, J., Clark, M. S., & Crockett, M. (2020). *How social relationships shape moral judgment*. PsyArXiv. https://doi.org/10.31234/osf.io/e7cgq

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does Rejection Hurt? An fMRI Study of Social Exclusion. *Science*, *302*, 4.

Eliasoph, N. (1998). *Avoiding politics: How Americans produce apathy in everyday life*. Cambridge University Press.

Ellison, N. B., & Boyd, D. M. (2013). *Sociality Through Social Network Sites* (W. H. Dutton, Ed.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199589074.013.0008

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Ellison, N. B., Vitak, J., Gray, R., & Lampe, C. (2014). Cultivating Social Resources on Social Network Sites: Facebook Relationship Maintenance Behaviors and Their Role in Social Capital Processes. *Journal of Computer-Mediated Communication*, *19*(4), 855–870. https://doi.org/10.1111/jcc4.12078

Englander, E. K., & Muldowney, A. M. (2007). *Just Turn the Darn Thing Off: Understanding Cyberbullying.* 11.

Epley, N., & Kruger, J. (2005). When what you type isn't what they read: The perseverance of stereotypes and expectancies over e-mail. *Journal of Experimental Social Psychology*, *41*(4), 414–422. https://doi.org/10.1016/j.jesp.2004.08.005

Erz, A., Marder, B., & Osadchaya, E. (2018). Hashtags: Motivational drivers, their use, and differences between influencers and followers. *Computers in Human Behavior*, *89*, 48–60. https://doi.org/10.1016/j.chb.2018.07.030

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature*, *13*(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, *370*(6516), 533–536. https://doi.org/10.1126/science.abe1715

Fisher, C. D. (1979). Transmission of positive and negative feedback to subordinates: A laboratory investigation. *Journal of Applied Psychology*, *64*(5), 533–540. https://doi.org/10.1037/0021-9010.64.5.533

Fox, J., & Moreland, J. J. (2015). The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in Human Behavior*, *45*, 168–176. https://doi.org/10.1016/j.chb.2014.11.083

García-Ramírez, G. M., Bogen, K. W., Rodríguez-Guzmán, V. M., Nugent, N., & Orchowski, L. M. (2019). #4645Boricuas: Twitter reactions to the estimates of deaths by Hurricane María in Puerto Rico. *Journal of Community Psychology*, 1–23. https://doi.org/10.1002/jcop.22295

Geer, J. G. (1988). What Do Open-Ended Questions Measure? *Public Opinion Quarterly*, *52*(3), 365. https://doi.org/10.1086/269113

Grieve, R., Indian, M., Witteveen, K., Anne Tolan, G., & Marrington, J. (2013). Face-to-face or Facebook: Can social connectedness be derived online? *Computers in Human Behavior*, *29*(3), 604–609. https://doi.org/10.1016/j.chb.2012.11.017

Hanschmidt, F., Linde, K., Hilbert, A., Heller, S. G. R.-, & Kersting, A. (2016). Abortion Stigma: A Systematic Review. *Perspectives on Sexual and Reproductive Health*, *48*(4), 169–177. https://doi.org/10.1363/48e8516

Haslam, N. (2006). Dehumanization: An Integrative Review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness Matters: A Theoretical and Empirical Review of Consequences and Mechanisms. *Annals of Behavioral Medicine*, *40*(2), 218–227. https://doi.org/10.1007/s12160-010-9210-8

Henrich, J., & Boyd, R. (2001). Why People Punish Defectors: Weak Conformist Transmission can Stabilize Costly Enforcement of Norms in Cooperative Dilemmas. *Journal of Theoretical Biology*, *208*(1), 79–89. https://doi.org/10.1006/jtbi.2000.2202

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, *345*(6202), 1340–1343. https://doi.org/10.1126/science.1251560

Hosseini, H., Kannan, S., Zhang, B., & Poovendran, R. (2017). Deceiving Google's Perspective API Built for Detecting Toxic Comments. *ArXiv:1702.08138 [Cs]*. http://arxiv.org/abs/1702.08138

Hunt, M. G., Marx, R., Lipson, C., & Young, J. (2018). No More FOMO: Limiting Social Media Decreases Loneliness and Depression. *Journal of Social and Clinical Psychology*, *37*(10), 751–768. https://doi.org/10.1521/jscp.2018.37.10.751

Joinson, A. N. (2001). Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. *European Journal of Social Psychology*, *31*(2), 177–192. https://doi.org/10.1002/ejsp.36

Jourard. (1964). *The transparent self: Self-disclosure and well-being* (Vol. 17). Van Nostrand.

Jourard, & Lasakow, P. (1958). Some factors in self-disclosure. *The Journal of Abnormal and Social Psychology*, *56*(1), 91–98. https://doi.org/10.1037/h0043357

Kahn, J. H., Achter, J. A., & Shambaugh, E. J. (2001). Client distress disclosure, characteristics at intake, and outcome in brief counseling. *Journal of Counseling Psychology*, *48*(2), 203–211. https://doi.org/10.1037/0022-0167.48.2.203

Kang, M. J., Rangel, A., Camus, M., & Camerer, C. F. (2011). Hypothetical and Real Choice Differentially Activate Common Valuation Areas. *Journal of Neuroscience*, *31*(2), 461–468. https://doi.org/10.1523/JNEUROSCI.1583-10.2011

Kjell, O. N. E., Kjell, K., Garcia, D., & Sikström, S. (2019). Semantic measures: Using natural language processing to measure, differentiate, and describe psychological constructs. *Psychological Methods*, *24*(1), 92–115. https://doi.org/10.1037/met0000191

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition and Emotion*, *26*(8), 1412–1427. https://doi.org/10.1080/02699931.2012.667392

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073–1137. https://doi.org/10.1037/a0035618

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., & Scherlis, W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well-being? *American Psychologist*, *53*(9), 1017–1031. https://doi.org/10.1037/0003-066X.53.9.1017

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537–567. http://dx.doi.org.proxy.lib.umich.edu/10.1146/annurev.psych.50.1.537

Kross, E., Verduyn, P., Boyer, M., Drake, B., Gainsburg, I., Vickers, B., Ybarra, O., & Jonides, J. (2018). Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion*. https://doi.org/10.1037/emo0000416

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., Shablack, H., Jonides, J., & Ybarra, O. (2013). Facebook Use Predicts Declines in Subjective Well-Being in Young Adults. *PLOS ONE*, *8*(8), e69841. https://doi.org/10.1371/journal.pone.0069841

Kruger, J., Epley, N., Parker, J., & Ng, Z.-W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology*, *89*(6), 925–936. https://doi.org/10.1037/0022-3514.89.6.925

Kumar, A., & Epley, N. (2020). It's surprisingly nice to hear you: Misunderstanding the impact of communication media can lead to suboptimal choices of how to connect with others. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000962

Lane, D. S., Kim, D. H., Lee, S. S., Weeks, B. E., & Kwak, N. (2017). From Online Disagreement to Offline Action: How Diverse Motivations for Using Social Media Can Increase Political Information Sharing and Catalyze Offline Political Participation. *Social Media + Society*, *3*(3), 2056305117716274. https://doi.org/10.1177/2056305117716274

Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, *28*(2), 434–443. https://doi.org/10.1016/j.chb.2011.10.014

Laurenceau, J. P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner

responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology*, *74*(5), 1238–1251. https://doi.org/10.1037//0022-3514.74.5.1238

Lee, K.-T., Noh, M.-J., & Koo, D.-M. (2013). Lonely People Are No Longer Lonely on Social Networking Sites: The Mediating Role of Self-Disclosure and Social Support. *Cyberpsychology, Behavior, and Social Networking*, *16*(6), 413–418. https://doi.org/10.1089/cyber.2012.0553

Leimeister, J. M., Krcmar, H., Köbler, F., Vetter, C., & Riedl, C. (2011). Social Connectedness on Facebook—An Explorative Study on Status Message Usage. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1953431

Ling, H., & Björling, E. (2020). Sharing Stress With a Robot: What Would a Robot Say? *Human-Machine Communication*, *1*, 133–158. https://doi.org/10.30658/hmc.1.8

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Luo, M., & Hancock, J. T. (2020). Self-disclosure and social media: Motivations, mechanisms and psychological well-being. *Current Opinion in Psychology*, *31*, 110–115. https://doi.org/10.1016/j.copsyc.2019.08.019

Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on Happiness: Well-being is Related to Having Less Small Talk and More Substantive Conversations. *Psychological Science*, *21*(4), 539–541. https://doi.org/10.1177/0956797610362675

Milgram, S., & Sabini, J. (1978). On maintaining urban norms: A field experiment in the subway. In S. Baum (Ed.), *Advances in environmental psychology* (Vol. 1, pp. 31–40).

Miller, D. T., & Norman, S. A. (1975). Actor-observer differences in perceptions of effective control. *Journal of Personality and Social Psychology*, *31*(3), 503–515. https://doi.org/10.1037/h0076485

Moberly, N. J., & Watkins, E. R. (2008). Ruminative self-focus, negative life events, and negative affect. *Behaviour Research and Therapy*, *46*(9), 1034–1039. https://doi.org/10.1016/j.brat.2008.06.004

Moore, M. J., Nakano, T., Enomoto, A., & Suda, T. (2012). Anonymity and roles associated with aggressive posts in an online forum. *Computers in Human Behavior*, *28*(3), 861–867. https://doi.org/10.1016/j.chb.2011.12.005

Moreno, M. A., Jelenchick, L. A., Egan, K. G., Cox, E., Young, H., Gannon, K. E., & Becker, T. (2011). Feeling bad on Facebook: Depression disclosures by college students on a social networking site. *Depression and Anxiety*, *28*(6), 447–455. https://doi.org/10.1002/da.20805

Müller, K., & Schwarz, C. (2018). Fanning the Flames of Hate: Social Media and Hate Crime. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3082972

Müller, K., & Schwarz, C. (2019). From Hashtag to Hate Crime: Twitter and Anti-Minority Sentiment. *SSRN Electronic Journal*, 110.

Murphy, S. C. (2017). A Hands-On Guide to Conducting Psychological Research on Twitter. *Social Psychological and Personality Science*, *8*(4), 396–412. https://doi.org/10.1177/1948550617697178

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network?: The structure of the twitter follow graph. *Proceedings of the 23rd International Conference on World Wide Web - WWW '14 Companion*, 493–498. https://doi.org/10.1145/2567948.2576939

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me?: Message content in social awareness streams. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work - CSCW '10*, 189. https://doi.org/10.1145/1718918.1718953

Nadkarni, A., & Hofmann, S. G. (2012). Why Do People Use Facebook? *Personality and Individual Differences*, *52*(3), 243–249. https://doi.org/10.1016/j.paid.2011.11.007

Noelle-Neumann, E. (1974). The Spiral of Silence A Theory of Public Opinion. *Journal of Communication*, *24*(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Norris, A., Bessett, D., Steinberg, J. R., Kavanaugh, M. L., De Zordo, S., & Becker, D. (2011). Abortion Stigma: A Reconceptualization of Constituents, Causes, and Consequences. *Women's Health Issues*, *21*(3, Supplement), S49–S54. https://doi.org/10.1016/j.whi.2011.02.010

Obadimu, A., Mead, E., Maleki, M., & Agarwal, N. (2020). *Developing an Epidemiological Model to Study Spread of Toxicity on YouTube*. https://doi.org/10.1007/978-3-030-61255-9_26

Park, J. H., Shin, J., & Fung, P. (2018). Reducing Gender Bias in Abusive Language Detection. *ArXiv:1808.07231 [Cs]*. http://arxiv.org/abs/1808.07231

Park, J., Lee, D. S., Shablack, H., Verduyn, P., Deldin, P., Ybarra, O., Jonides, J., & Kross, E. (2016). When perceptions defy reality: The relationships between depression and actual

and perceived Facebook social support. *Journal of Affective Disorders*, *200*, 37–44. https://doi.org/10.1016/j.jad.2016.01.048

Park, N., Jin, B., & Annie Jin, S.-A. (2011). Effects of self-disclosure on relational intimacy in Facebook. *Computers in Human Behavior*, *27*(5), 1974–1983. https://doi.org/10.1016/j.chb.2011.05.004

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being Immersed in Social Networking Environment: Facebook Groups, Uses and Gratifications, and Social Outcomes. *CyberPsychology & Behavior*, *12*(6), 729–733. https://doi.org/10.1089/cpb.2009.0003

Patti M. Valkenburg, & Jochen Peter. (2009). Social Consequences of the Internet for Adolescents: A Decade of Research. *Current Directions in Psychological Science*, *18*(1), 1–5.

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing. *Journal of Personality and Social Psychology*, 1197–1208.

Planalp, S. (1993). Friends' and Acquaintances' Conversations II: Coded Differences. *Journal of Social and Personal Relationships*, *10*(3), 339–354. https://doi.org/10.1177/0265407593103003

Planalp, S., & Benson, A. (1992). Friends' and Acquaintances' Conversations I: Perceived Differences. *Journal of Social and Personal Relationships*, *9*(4), 483–506. https://doi.org/10.1177/0265407592094002

Qiu, L., Lin, H., Leung, A. K., & Tov, W. (2012). Putting Their Best Foot Forward: Emotional Disclosure on Facebook. *Cyberpsychology, Behavior, and Social Networking*, *15*(10), 569–572. https://doi.org/10.1089/cyber.2012.0200

Rauschnabel, P. A., Sheldon, P., & Herzfeldt, E. (2019). What motivates users to hashtag on social media? *Psychology & Marketing*, *36*(5), 473–488. https://doi.org/10.1002/mar.21191

Reis, H. T., & Shaver, P. (1988). Intimacy as an interpersonal process. In *Handbook of Personal Relationships: Theory, Research, and Interventions*.

Rimé, B. (2009). Emotion Elicits the Social Sharing of Emotion: Theory and Empirical Review. *Emotion Review*, *1*(1), 60–85. https://doi.org/10.1177/1754073908097189

Rimé, B., Mesquita, B., Boca, S., & Philippot, P. (1991). Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition & Emotion*, *5*(5–6), 435–465. https://doi.org/10.1080/02699939108411052

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, *68*(3), 404–417. https://doi.org/10.1037/0022-3514.68.3.404

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rose, A. J. (2002). Co-Rumination in the Friendships of Girls and Boys. *Child Development*, *73*(6), 1830–1843.

Rosen, S., & Tesser, A. (1970). On Reluctance to Communicate Undesirable Information: The MUM Effect. *Sociometry*, *33*(3), 253–263. https://doi.org/10.2307/2786156

Rösner, L., & Krämer, N. C. (2016). Verbal Venting in the Social Web: Effects of Anonymity and Group Norms on Aggressive Language Use in Online Comments. *Social Media + Society*, *2*(3), 205630511666422. https://doi.org/10.1177/2056305116664220

Rusbult, C. E., & Buunk, B. P. (1993). Commitment Processes in Close Relationships: An Interdependence Analysis. *Journal of Social and Personal Relationships*, *10*(2), 175–204. https://doi.org/10.1177/026540759301000202

Ryan, R. M., & Deci, E. L. (2001). On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being. *Annual Review of Psychology*, *52*(1), 141–166. https://doi.org/10.1146/annurev.psych.52.1.141

Sabini, J., Siepmann, M., Stein, J., & Meyerowitz, M. (2000). Who is Embarrassed by What? *Cognition & Emotion*, *14*(2), 213–240. https://doi.org/10.1080/026999300378941

Sandstrom, G. M., & Dunn, E. W. (2014). Is Efficiency Overrated?: Minimal Social Interactions Lead to Belonging and Positive Affect. *Social Psychological and Personality Science*, *5*(4), 437–442. https://doi.org/10.1177/1948550613502990

Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology*, *47*(1), 269–273. https://doi.org/10.1016/j.jesp.2010.09.005

Sawaoka, T., & Monin, B. (2018). The Paradox of Viral Outrage. *Psychological Science*, 095679761878065. https://doi.org/10.1177/0956797618780658

Schein, C., & Gray, K. (2016). Moralization and Harmification: The Dyadic Loop Explains How the Innocuous Becomes Harmful and Wrong. *Psychological Inquiry*, *27*(1), 62–65. https://doi.org/10.1080/1047840X.2016.1111121

Schein, C., & Gray, K. (2018). The Theory of Dyadic Morality: Reinventing Moral Judgment by Redefining Harm. *Personality and Social Psychology Review*, *22*(1), 32–70. https://doi.org/10.1177/1088868317698288

Schlenker, B. R., & Leary, M. R. (1982). Audiences' reactions to self-enhancing, self-denigrating, and accurate self-presentations. *Journal of Experimental Social Psychology*, *18*(1), 89–104. https://doi.org/10.1016/0022-1031(82)90083-X

Schroeder, J., & Epley, N. (2015). The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. *Psychological Science*, *26*(6), 877–891. https://doi.org/10.1177/0956797615572906

Schroeder, J., Kardas, M., & Epley, N. (2017). The Humanizing Voice: Speech Reveals, and Text Conceals, a More Thoughtful Mind in the Midst of Disagreement. *Psychological Science*, *28*(12), 1745–1762. https://doi.org/10.1177/0956797617713798

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8–20. https://doi.org/10.1016/j.pragma.2015.03.015

Shakya, H. B., & Christakis, N. A. (2017). Association of Facebook Use With Compromised Well-Being: A Longitudinal Study. *American Journal of Epidemiology*. https://doi.org/10.1093/aje/kww189

Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The Power of the *Like* in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media. *Psychological Science*, *27*(7), 1027–1035. https://doi.org/10.1177/0956797616645673

Smith, A. (2011, November 15). Why Americans use social media. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2011/11/15/why-americans-use-social-media/

Sprecher, S., Treger, S., & Wondra, J. D. (2013). Effects of self-disclosure role on liking, closeness, and other impressions in get-acquainted interactions. *Journal of Social and Personal Relationships*, *30*(4), 497–514. https://doi.org/10.1177/0265407512459033

Suler, J. (2004). The Online Disinhibition Effect. *CyberPsychology & Behavior*, *7*(3).

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences*, *109*(21), 8038–8043. https://doi.org/10.1073/pnas.1202129109

Taylor, S. E., & Armor, D. A. (1996). Positive Illusions and Coping with Adversity. *Journal of Personality*, *64*(4), 873–898. https://doi.org/10.1111/j.1467-6494.1996.tb00947.x

Tedeschi, J. T. (1986). Private and Public Experiences and the Self. In R. F. Baumeister (Ed.), *Public Self and Private Self* (pp. 1–20). Springer. https://doi.org/10.1007/978-1-4613-9564-5_1

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, *7*(7), 320–324. https://doi.org/10.1016/S1364-6613(03)00135-9

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870. https://doi.org/10.1037/0022-3514.78.5.853

Tice, D. M., Butler, J. L., Muraven, M. B., & Stillwell, A. M. (1995). When modesty prevails: Differential favorability of self-presentation to friends and strangers. *Journal of Personality and Social Psychology*, *69*(6), 1120–1138. https://doi.org/10.1037/0022-3514.69.6.1120

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, *26*(3), 277–287. https://doi.org/10.1016/j.chb.2009.11.014

Toma, C. L., & Hancock, J. T. (2013). Self-Affirmation Underlies Facebook Use. *Personality and Social Psychology Bulletin*, *39*(3), 321–331. https://doi.org/10.1177/0146167212474694

Toma, C. L., Hancock, J. T., & Ellison, N. B. (2008). Separating Fact From Fiction: An Examination of Deceptive Self-Presentation in Online Dating Profiles. *Personality and Social Psychology Bulletin*, *34*(8), 1023–1036. https://doi.org/10.1177/0146167208318067

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Tur, B., Harstad, J., & Antonakis, J. (2021). Effect of charismatic signaling in social media settings: Evidence from TED and Twitter. *The Leadership Quarterly*, 101476. https://doi.org/10.1016/j.leaqua.2020.101476

Utz, S. (2015). The function of self-disclosure on social network sites: Not only intimate, but also positive and entertaining self-disclosures increase the feeling of connection. *Computers in Human Behavior*, *45*, 1–10. https://doi.org/10.1016/j.chb.2014.11.076

Valenzuela, S., Park, N., & Kee, K. F. (2009). Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation. *Journal of Computer-Mediated Communication*, *14*(4), 875–901. https://doi.org/10.1111/j.1083-6101.2009.01474.x

van der Heijden, H. (2004). User Acceptance of Hedonic Information Systems. *MIS Quarterly*, *28*(4), 695–704. https://doi.org/10.2307/25148660

Verduyn, P., Lee, D. S., Park, J., Shablack, H., Orvell, A., Bayer, J., Ybarra, O., Jonides, J., & Kross, E. (2015). Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence. *Journal of Experimental Psychology: General*, *144*(2), 480–488. https://doi.org/10.1037/xge0000057

Vittengl, J. R., & Holt, C. S. (2000). Getting Acquainted: The Relationship of Self-Disclosure and Social Attraction to Positive Affect. *Journal of Social and Personal Relationships*, *17*(1), 53–66. https://doi.org/10.1177/0265407500171003

Walsh, R. M., Forest, A. L., & Orehek, E. (2020). Self-disclosure on social media: The role of perceived network responsiveness. *Computers in Human Behavior*, *104*, 106162. https://doi.org/10.1016/j.chb.2019.106162

Walther, J. B. (2007). Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, *23*(5), 2538–2557. https://doi.org/10.1016/j.chb.2006.05.002

Walther, J. B., Van Der Heide, B., Ramirez, A., Burgoon, J. K., & Peña, J. (2015). Interpersonal and Hyperpersonal Dimensions of Computer-Mediated Communication. In S. S. Sundar (Ed.), *The Handbook of the Psychology of Communication Technology* (pp. 1–22). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118426456.ch1

Wang, Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., & Cranor, L. F. (2011). *"I regretted the minute I pressed share": A qualitative study of regrets on Facebook*. 1. https://doi.org/10.1145/2078827.2078841

Watson, D., Clark, L. A., McIntyre, C. W., & Hamaker, S. (1992). Affect, personality, and social activity. *Journal of Personality and Social Psychology*, *63*(6), 1011–1025. https://doi.org/10.1037/0022-3514.63.6.1011

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033. https://doi.org/10.1016/j.jesp.2013.07.002

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2019). Punish or Protect? How Close Relationships Shape Responses to Moral Violations. *Personality and Social Psychology Bulletin*, 0146167219873485. https://doi.org/10.1177/0146167219873485

Whiting, A., & Williams, D. (2013). Why people use social media: A uses and gratifications approach. *Qualitative Market Research: An International Journal*, *16*(4), 362–369. https://doi.org/10.1108/QMR-06-2013-0041

Whittaker, E., & Kowalski, R. M. (2015). Cyberbullying Via Social Media. *Journal of School Violence*, *14*(1), 11–29. https://doi.org/10.1080/15388220.2014.949377

Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2019). Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. *The British Journal of Criminology*, azz049. https://doi.org/10.1093/bjc/azz049

Wood, J. V., & Forest, A. L. (2016). Chapter Three - Self-Protective yet Self-Defeating: The Paradox of Low Self-Esteem People's Self-Disclosures. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 53, pp. 131–188). Academic Press. https://doi.org/10.1016/bs.aesp.2015.10.001

Zaki, J., & Williams, W. C. (2013). Interpersonal emotion regulation. *Emotion*, *13*(5), 803–810. https://doi.org/10.1037/a0033839

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, *25*(3), 274–291. https://doi.org/10.1080/10350330.2014.996948

Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239